10TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

PART B

Edited by
R. M. BRITO ALVES,
C. A. OLLER NASCIMENTO,
E. C. BISCAIA Jr

COMPUTER-AIDED CHEMICAL ENGINEERING, 27

10th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

COMPUTER-AIDED CHEMICAL ENGINEERING

Advisory Editor: R. Gani

10th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING: PART B

Edited by

Rita Maria de Brito Alves *and* Claudio Augusto Oller do Nascimento
University of São Paulo

Brazil

Evaristo Chalbaud Biscaia Jr.
Federal University of Rio de Janeiro
Rio de Janeiro
RJ
Brazil

For information on all Elsevier publications
visit our web site at books.elsevier.com

Part B

Operations and Control

Oral

Operations and Control

Poster

Non-traditional Applications of Process Systems Engineering

Oral

Non-traditional Applications of Process Systems Engineering

Poster

Business Decision Support

Oral

Business Decision Support

Poster

Education in Process Systems Engineering

Oral

Education in Process Systems Engineering

Poster

On the Topological Analysis of Industrial Process Data Using the SOM

Francesco Corona,[a] Michela Mulas,[b] Roberto Baratti,[c] Jose Romagnoli[c,*]

[a]*Dept. of Information and Computer Science, Helsinki University of Technology TKK, P.O. Box 5400, Espoo FI-02015 TKK, Finland*
[b]*Dept. of Biotechnology and Chemical Technology, Helsinki University of Technology TKK, P.O. Box 6100, Espoo FI-02015 TKK, Finland*
[c]*Dept. of Chemical Engineering and Materials, University of Cagliari, Piazza d'Armi, Cagliari I-09123, Italy*

Abstract

In this paper, we overview and discuss the implementation of some topological approaches to modeling and analyzing industrial process data. The discussed methods are used in visualizing process measurements and extracting information by exploiting the metric structure of the observations. Emphasis is given to modeling with the Self-Organizing Map (SOM). The SOM is a standard method for dimensionality reduction and vector quantization equipped with many displays for visualization. Some of the possibilities of the SOM with process data are discussed by exploring measurements from a full-scale gas treatment plant where the goal is to identify important operational modes and sensitive process variables before developing an alternative control strategy.

Keywords: Process monitoring, Process supervision, the Self-Organizing Map.

1. Introduction

Modern process plants are motivated to maintaining and improving product quality and profit while satisfying more stringent environmental and safety constraints. For efficient operation, any decision-making action related to plant operation requires the knowledge of the actual state of the processes. The availability of easily accessible displays and intuitive knowledge of the states is invaluable with immediate implications for profitability, management planning, environmental responsibility and safety.

Due to the advances in measurement and information technology, process industry is also passing through the era of data explosion; historical data are available in abundance. The data derived approach to process monitoring and supervision seeks to construct a representation of a process from a set of easily accessible measurements that quantify its behavior, without explicitly assuming any *a priori* knowledge of the underlying phenomena. However, the task is daunting and remains a major concern. It is necessary the availability of effective methods that *i*) model the data to extract the structures existing in the measurements, *ii*) automatically identify and reconstruct the most relevant structures for the scope at hand, and *iii*) allow for easily interpretable displays where the states' information is presented to the plant operators. Intuitive knowledge of all visited states is invaluable for a safe plant operation and trustworthy methods are a necessary part in a modern supervision and control strategy.

---

[*] On leave, from the Department of Chemical Engineering, Louisiana State University, USA.

In this paper, we discuss a data derived approach to model, visualize and analyze the information encoded in industrial process data. The approach is based on a classical machine learning method, the Self-Organizing Map, SOM (Kohonen, 2001). The SOM combines many of the properties of other techniques and shares many commonalities with two standard methods for data projection (the Principal Components Analysis, PCA; Jolliffe, 2002) and clustering (the K-means; Hartigan and Wong, 1979). In addition, the SOM is also provided with a set of tools that allow for efficient data visualization in high-dimensional settings. Kaski (1997) and Vesanto (2002) discussed the use of the Self-Organizing Map in exploratory data analysis, and their methods are widely known in many fields, see Oja et al. (2003). The major contributions in using the SOM on process data were proposed by Alhoniemi (2002) and Laine (2003). The SOM is here used as a common framework to identify the process modes for a full-scale gas treatment plant and present the extracted information on intelligible displays.

2. Topological data analysis with the SOM

The Self-Organizing Map is an adaptive formulation of the quantization paradigm on a low dimensional space. Thus, the method singles out from other machine learning techniques because it performs both data clustering and projection. The main advantage of the SOM is in the visualization tools that allow for an intuitive data exploration.

The basic SOM consists of a low-dimensional regular array of K nodes where a prototype vector $\mathbf{m}_k \in \mathbf{R}^p$ is associated with every node k. Each prototype acts as an adaptive model for the observations $\mathbf{v}_i \in \mathbf{R}^p$. During the computation, the data are mapped onto the SOM array and the model vectors adapted according to the expression:

$$\mathbf{m}_k(t+1) = \mathbf{m}_k(t) + \alpha(t)h_{k,c(\mathbf{v}_i)}(\mathbf{v}_i(t) - \mathbf{m}_k(t)). \tag{1}$$

In the learning rule in Eq. 1, t denotes the discrete coordinate of the mapping steps and $\alpha(t) \in (0,1)$ is the decreasing learning rate. The map is computed recursively for each observation. The scalar multiplier $h_{k,c(vi)}$ indicates the neighborhood kernel that, if chosen in its Gaussian form, as in Eq. 2, acts as a smoothing function

$$c(t) = \alpha(t)\exp\left(-\left\|\mathbf{r}_l - \mathbf{r}_c\right\|^2 / 2\sigma^2(t)\right), \tag{2}$$

centered at the Best Matching Unit, BMU (the model $\mathbf{m}_c(t)$ that best matches with the observation $\mathbf{v}_i$). The vectors $\mathbf{r}_l$ and $\mathbf{r}_c$ (both in $\mathbf{R}^2$, for the 2D map) represent the geometric location of the nodes on the array. Upon performing a parallel comparison algorithm, the subscript c (in Eq. 1 and 2) is assigned to the BMU. The competitive criterion for comparison is usually the Euclidean norm. During calibration, the model vectors learn a nonlinear manifold in the original space such that the relevant topological and metric properties of the observations are preserved on the low-dimensional map. Here, the structures are represented by geometric relationships.

In the case of a slightly modified algorithm with a fixed kernel function, the SOM can be also understood from the optimization of a cost function (Heskes, 1997). The cost closely relates to the objective of the K-means algorithm; the only difference being in the neighbourhood function. In that sense, the SOM operates as the conventional clustering method, if the width of the kernel is constantly set to zero. The SOM has also neat projection properties: the cost also closely resembles the objective optimized by Curvilinear Components Analysis, CCA (Demartines and Herault, 1997).

In the typical case of projections onto 2D arrays, the SOM offers excellent displays for data exploration. In that sense, the approach to topological data analysis with the SOM is mainly visual and intuitive. The visualization techniques used in the experiments are the common component planes and the U-matrix (Vesanto, 2002).

3. Application to an industrial deethanizer

To illustrate the potentialities of topological data analysis using the SOM, a set of measurements from a full-scale process was considered. The monitoring problem consists of modeling and analyzing the operational behavior of an industrial deethanizer, starting from a set of online process measurements. The motivation for choosing this unit is merely illustrative. The column offers a variety of behaviors that reflects the operational usage, hence an interesting groundwork for discussion.

The deethanizer, in Fig. 1, is operated to separate ethane from a feed (a light naphtha) while minimizing the ethane extracted from the bottom (an economical constraint for the subsequent unit in the plant). The operational constraint is in the maximum ethane loss from the bottom; at present, the operational threshold is set to be smaller than 2%.
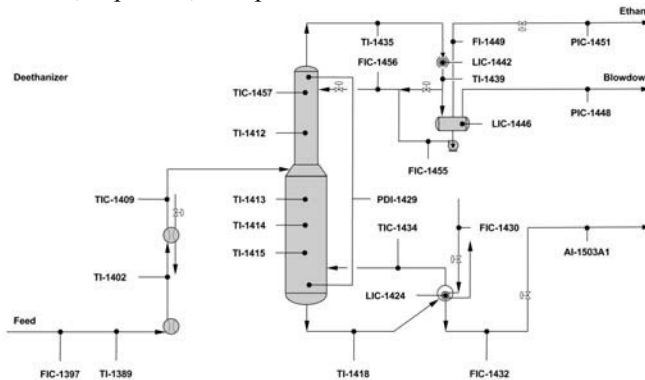
Figure 1. Deethanizer: Simplified flow-sheet.

In order to analyze the behavior of the unit, measurements from a set of process variables was collected from the plant's distributed control system (DCS). The data correspond to 3 weeks of continuous operation in winter asset and 3 weeks in summer asset and are available as 3-minute averages; overall, 27 variables are acquired, for a macroscopic characterization of the unit. In addition, there are several control loops in the process. The column temperature TIC-1457 and the vapour temperature TIC-1434 out of the reboiler is controlled by manipulating the reflux flow-rate FIC-1456 and the steam flow-rate FIC-1430 to the reboiler, respectively; both loops are cascaded to the corresponding flow-rates. The distillate pressure PIC-1451 is controlled by the distillate flow-rate FI-1449 and the reboiler's level LIC-1424 by the bottom flow-rate FIC-1432.

The operational objective of the column is to produce as much ethane as possible (minimizing the propane's concentration from the top of the column) while satisfying the constraint on the impurity from the bottom (ethane's concentration smaller than 2%). Based on the ethane loss from the bottom, three operational modes are identified:

- a *normal* status, corresponding to the operation of the column, where the concentration of ethane is within allowable bounds (in the 1.8% - 2.0% range),
- a *high* status, corresponding to the operation of the column, where the concentration of ethane is exceeding the allowable upper bound (>2%),
- a *low* status, corresponding to the operation of the column, where the concentration of ethane is below the allowable lower bound (<1.8%).

The two abnormal conditions have direct economic implications; when at *low* status, the process is delivering a product out of specifications whereas, when at *high* status the product is within the specifications, but an unnecessary operational cost is observed.

3.1. Exploration and analysis

To understand under which conditions such modes are experienced, the data were analyzed by exploring their clustering structure. Starting from a selection of relevant variables, we augmented this subset with an additional *dummy* indicator, specifically calculated to indicate the status. The new variable was defined so as to take values +1, -1 and 0, according to the operational status. Value 0 is assigned to the *normal* mode, whereas +1 and 1 correspond to *high* and *low* modes, respectively. Notice that the *dummy* variable requires online measurements for the ethane concentration; such a variable, (AI-1503A1), is available in the DCS from a continuous-flow chromatograph.

The set of variables, augmented by the *dummy* indicator, was used to learn a 2D SOM, from the winter observations only. The map consists of a hexagonal array of prototypes initialized in the space spanned by the eigenvectors corresponding to the two largest eigenvalues of the covariance matrix of the data. As usual, the ratio between the two largest eigenvalues was used to calculate the ratio between the two dimensions of the map. On the resulting map, the SOM's component planes and U-matrix were analyzed.
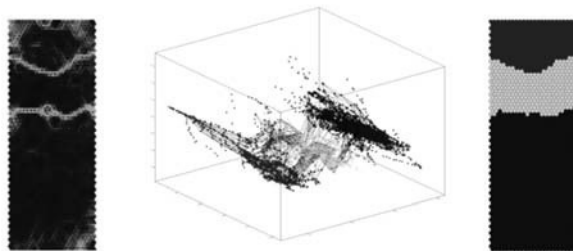
Figure 2. The U-Matrix with three clearly separated clusters in blue (left), the clustered SOM and the winter data after projection onto the principal components space (center) and the map dyed according to the cluster membership obtained from the K-means algorithm (right).

The U-matrix is the display used to visualize the distances between each prototype and its neighbors. The common way to present it consists of an initial projection of the distances onto a color axis and the subsequent display with dyed markers between each node. In the U-matrix, areas with homogeneous coloring (corresponding to small within-cluster distances) are recognized as clusters, whereas cluster borders are identified still as areas with homogeneous coloring but corresponding to large between-cluster distances. The U-matrix is shown in Fig. 2 (left), with a color-coding that assigns blue to close together prototypes (clusters) fading toward red as the distance increases (cluster separations). The visualization permits to recognize 3 distinct taxonomies, as well as several minor substructures. An analogous visualization of the prototypes' grouping is achieved by projecting the nodes onto a low-dimensional subspace, a 3D PCA space (in Fig. 2 (center)). Indeed, also this visualization illustrates the clustering. However, to obtain a quantitative characterization, the prototypes of the SOM can be regarded as a reduced data set and modeled separately with a specific clustering algorithm. We adopted the K-means coupled by the Davier-Bouldin index (as measure of cluster validity) to identify the optimal number K of taxonomies (Milligan and Cooper, 1995). As expected, optimality was found for $K=3$. The operational modes are located in the lower, middle and upper part of the map. The mode-cluster association is straightforward after dying the SOM's region in Fig. 2 (right) with different membership colors (blue, green and red) from the K-means according to a scheme that assigns them to the operational modes (+1, 0 and -1, respectively), and comparing it with the component planes of the *dummy* and ethane (Fig. 3 (left) and (center)).
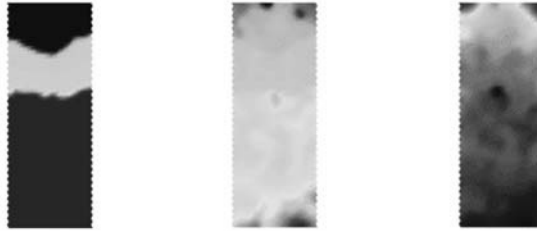
Figure 3. The component planes for the *dummy* variable (left), the ethane concentration (center) and the temperature TI-1414. The coloring ranges from blue (low values) to red (high values).

Fig. 3 also depicts the plane of the column temperature TI-1414 (right); looking for similar patterns in similar positions in the components planes shows a neat dependence between ethane concentrations and temperature measurements. Such pair of variables is characterized by near identical but reversed component planes, thus highlighting the presence of an inherent inverse correlation between the two variables.

To maintain the ethane concentration in normal conditions (within the 1.8-2.2% range, Fig. 4 (left)), the analysis suggests that such temperature should be controlled (within the 52-55°C range, Fig. 4 (center)). The information can be easily retrieved after dyeing the point measurements of the two variables with the corresponding mode membership. A possible variable to manipulate is the steam flow-rate FIC-1430, such a variable is, however, presently unused for control. During the winter period, an overall 85% of off-spec operation, was in fact observed, Fig. 4 (right). The summary is obtained by calculating the number of point measurements falling outside the normality conditions over the total number of observations, and it can be also visually recovered from the area distribution of the clusters in the SOM depicted in Fig 2 (right).
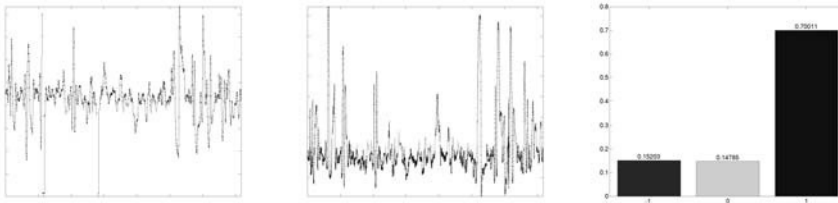
Figure 4. The colored time series for AI-1503A1 (left), TI-1414 (center) and the modes distribution during winter operation (right).

The winter map can also be used to visualize the behavior of the unit in extrapolation; for example, on the summer measurements. The analysis initially projects the new data onto the calibrated SOM, being the mapping based on a nearest neighbor criterion between the new observations and the prototypes. The winter map can be then enhanced by the inclusion of the summer point trajectories followed by the unit. The trajectory, which passes through all the BMUs of each new set of measurements, is shown as red line connecting the visited prototypes, marked as yellow dots. The trajectory makes it possible to indicate the current mode of the process and observe how it was reached.

In Fig. 5, the process trajectory is reported for a small time window corresponding to fifteen hours of continuous summer operation of the deethanizer. Following the temporal evolution, the diagrams show a process that is initially operated in *normal* condition, as for the ethane in the bottom and reference temperature. As the process moved further in time, new prototype vectors were visited and added to the trajectory

until the column leaves the normality region and crosses it towards the region of low ethane composition and starts oscillating. In a similar fashion, all the process variables changed coloring to match the visited modes. The abnormal situation was mainly due to an abrupt decrease in feed flow-rate and possibly composition. In turns, the variation triggered the action on steam to reboiler flow-rate FIC-1430, as well as the reflux to control the top temperature FIC-1436.
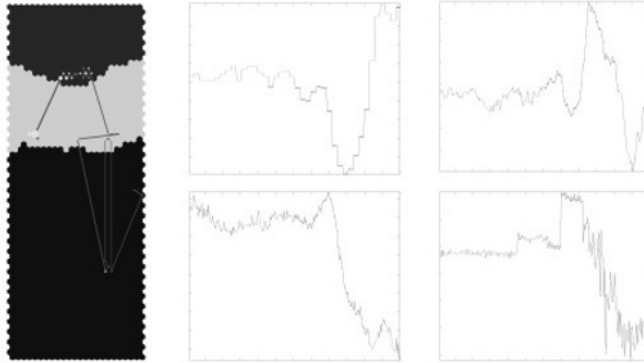
Figure 5. The trajectory and the colored time evolution for the ethane AI-1502A1, the temperature TI-1414, the steam to reboiler FIC-1430 and the reflux flow-rate FIC-1456, in clockwise order

4. Conclusions

In this work, we implemented and discussed a strategy to model, visualize and analyze the information encoded in industrial process data using the Self-Organizing-Map. In particular, the proposed strategy was applied to an industrial distillation column allowing us to illustrate on simple displays how the clustering structure of the measurements corresponds to the operational modes of the units and the possibility to define an alternative control strategy for maintaining it in the normal operating state.

References

E. Alhoniemi, 2002, Unsupervised pattern recognition methods for exploratory analysis of industrial process data, Ph.D. dissertation. Helsini University of Technology, Finland.

P. Demartines, J. Herault, 1997, Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets, IEEE Trans. on Neural Networks 8, 148–154.

A. Hartigan, M.A. Wong, 1979, A K-means clustering algorithm. Applied Statistics 28, 100–108.

T. Heskes, 1999, Energy functions for self-organizing maps, in Kohonen maps, 303-316.

I. T. Jolliffe, 2002, Principal Components Analysis: Second Edition, Springer.

S. Kaski, 1997, Data exploration using self-organizing maps, Ph.D. dissertation ISBN 952-5148-13-0. Helsinki University of Technology, Finland.

T. Kohonen, 2001 , Self organizing maps: Second Edition, Springer.

S. Laine, 2003, Variable selection and feature extraction to learn from industrial data. Ph.D. dissertation ISBN 951-22-6669-5, Helsinki University of Technology, Finland.

G.W. Milligan, M.C. Cooper, 1995, An examination of procedures for determining the number of clusters in a dataset, Psychometrika 50, 159–179.

M. Oja, S. Kaski, T. Kohonen, 2003, Bibliography of Self-Organizing Map (SOM) papers: 1998-2001 addendum, Neural Computing Surveys 3, 1–156.

J. Vesanto, 2002, Data exploration based on the Self-Organizing Map, Ph.D. dissertation, Helsinki University of Technology, Finland.

Delaunay Tessellation and Topological Regression: An Application to Estimating Product Properties from Spectroscopic Measurements

Francesco Corona,[a] Elia Liitiäinen,[a] Amaury Lendasse,[a] Roberto Baratti,[b]

Lorenzo Sassu[c]

[a]*Dept. of Information and Computer Science, Helsinki University of Technology TKK, P.O. Box 5400, Espoo FI-02015 TKK, Finland*
[b]*Dept. of Chemical Engineering and Materials, University of Cagliari, Piazza d'Armi, Cagliari I-09123, Italy*
[c]*Process Department., Sartec - Saras Ricerche e Tecnologie S.p.A. Strada 5 Traversa C, Assemini I-09032, Italy*

Abstract

The Delaunay tessellation and topological regression is a local simplex method for multivariate calibration. The method, developed within computational geometry has potential for applications in analytical chemistry and process monitoring. This study investigates the applicability of the method for estimating the aromatic composition in Light Cycle Oil (LCO) by Near Infrared (NIR) spectroscopy.

Keywords: Process monitoring, multivariate calibration, Delaunay tessellation, nearest-neighbor regression.

1. Introduction

Real-time monitoring is an essential component in modern process industry for optimizing production toward high-quality products while reducing operating and off-specification costs. The tools of process analytical chemistry like Infrared (IR) and Near Infrared (NIR) spectroscopy fulfill the necessary requirements for real-time analysis of important properties for a broad variety of materials (Workman, 1999).

The principle underlying process monitoring from spectra is the existence of a relationship between the spectrum of a given product and the property of interest. The relationship is rarely known *a priori* but it can be reconstructed from data by calibrating multivariate models. Multivariate calibration methods are often divided into local and global approaches. The latter use all known (calibration) observations to learn the parameters of a single regression model. The former use only small subsets of the calibration data to build different calibration models located in the neighborhood of the observation whose properties have to be estimated. Classical parametric approaches like Principal Components Regression (PCR) and Partial Least Squares Regression (PLSR) are available in both local and global variants (Gemperline *et al*., 2006). Among local methods, non-parametric approaches as topological regression (Jin *et al*., 2003a and Jin *et al*., 2003b) have gained recent interest, mostly driven by industrial motivations; e.g., an inherent ability to handle nonlinearities and the possibility to minimizing models' maintenance tasks while retaining the prediction accuracy.

With the goal to investigate alternative calibration methods that reduce the maintenance costs associated to continuous recalibrations, this paper discusses the Delaunay

Tessellation and Topological Regression (DTR) method (Jin *et al*., 2003b) and its application to calibrating the aromatic composition in Light Cycle Oil (LCO) by NIR spectroscopy. The application is framed within the intense research activity that has characterized the recent trends in refining aimed at optimizing the use of low-value products. LCO is a low-value stream in the diesel boiling range produced in Fluid Catalytic Cracking (FCC) units. Due to its poor characteristics (i.e., high total aromatic content and a considerable percentage of compact structure poliaromatics, among others), LCO is usually not blended directly in the finished diesel fuel pool but it can be upgraded to the higher value diesel if processed in hydrotreatment units (where aromatic compounds are hydrogenated). Within this framework and in order to assure the required process and environmental standards of the hydrotreated products, rapid and cost effective (and possibly on-line) evaluation of the aromatic content is mandatory. Among the many chemometric methods available for such purposes, the DTR method has recently been proposed as an application of computational geometry. It singles out from standard approaches because of its potentiality to achieve accuracies comparable with PCR and PLSR while being much simpler to develop and maintain/upgrade (Jin *et al*., 2005 and Jin *et al*., 2006). In order to assess the potential of the method, a feasibility study with comparison to standard regression methods was performed.

2. Delaunay tessellation and regression

The Delaunay Tessellation and Topological Regression is a local multivariate calibration method developed from arguments in computational geometry. In its basic formulation, the DTR consists of three main steps: 1) a dimensionality reduction based on the known input observations (e.g., NIR spectra), 2) the generation (in the low-dimensional space) of an unstructured mesh by Delaunay tessellation and 3) a nearest-neighbor (or topological) regression for estimating the outputs (e.g., the aromatic content in hydrocarbon mixtures) for a set of unknown observations.

2.1. Dimensionality reduction
The dimensionality reduction step aims at projecting the input observations onto a system of lower coordinates in such a way that certain properties of the data points are preserved as faithfully as possible. The mapping can be either driven only by the input data points (as in PCR) or by both the input and output data points (as in PLSR).
In general, there is a wide range of other methods to performing dimensionality reduction (e.g., see Lee and Verleysen, 2007) that can be considered in this phase. This study is confined to a projection based only on input data, because this representation can be common to all the output properties; thus, minimizing problems and costs associated to the models' recalibration and maintenance. For the sake of simplicity, a PCA was used to characterize the experiments; in that sense, the local property of the data that is preserved is in the set of pair-wise distances between them (Jolliffe, 2002).

2.2. Delaunay tessellation
Once the input observations are projected onto a low dimensional system of coordinates (e.g., the principal components), the known part of this space is partitioned, by generating a mesh using all the available data points. The elements of the mesh are simplexes delimited by known observations (i.e., input data points with known values for the output properties, the calibration set). Within each simplex, the locality conditions are assumed because similar data should be mapped close to each other.
A well-known method for generating a mesh of simplexes is the Delaunay tessellation (Gudmundsson *et al*. 2002). For a given set of point observations in two dimensions, the Delaunay tessellation constructs a mesh of triangular simplexes (hence, the common

name Delaunay *triangularization*) by using all the input data points as vertices; one triangle is a valid simplex if and only if its circumcircle does not enclose any other point in the set (the empty circle condition). In 3 dimensions, the simplexes are tetrahedrons, and so on. Generally, the Delaunay tessellation of a point set exists and is unique.

Notice that the DTR method requires a mesh generation performed on as many dimensions as those obtained in the dimensionality reduction step. For a reduction to *s* dimensions, the elements of the tessellation are simplexes defined by *k=s+1* points.

2.3. Topological regression

Once the mesh is built, it is used for estimating the properties of new observations; that is, data points for which only the input values are known. Also the new observations must be projected onto the same low dimensional system obtained in the first step.

The standard case for estimation is when a new observation **x**(i) (in the *s*-dimensional projection space) happens to fall within the convex hull that contains the know data points. In such a situation, the new point also falls within one of the simplexes.

In two dimensions, the simplex is a triangle with vertexes **x**(1), **x**(2) and **x**(3) (three known observations, also in the *s*-dimensional space), and the position of the new observation with respect to its three neighboring points (the vertexes) is found as a weighted sum of their input coordinates subjected to the convexity constraints (i.e., the weights, **α**(i), are bounded to the unit interval and sum up to one). For each such new data point, only the enclosing triangular simplex fulfills the convexity constraints.

The following equations are used for calculating **α**(i) in two dimensions:

$$\alpha_1(i) = \frac{(x_2(i) - x_2(2))(x_1(3) - x_1(2)) + (x_1(i) - x_1(2))(x_2(2) - x_2(3))}{(x_2(1) - x_2(2))(x_1(3) - x_1(2)) + (x_1(1) - x_1(2))(x_2(2) - x_2(3))},$$

$$\alpha_2(i) = \frac{(x_2(i) - x_2(1))(x_1(3) - x_1(1)) + (x_1(i) - x_1(1))(x_2(1) - x_2(3))}{(x_2(2) - x_2(1))(x_1(3) - x_1(1)) + (x_1(2) - x_1(1))(x_2(1) - x_2(3))}, \tag{1}$$

$$\alpha_3(i) = \frac{(x_2(i) - x_2(2))(x_1(1) - x_1(2)) + (x_1(i) - x_1(2))(x_2(2) - x_2(1))}{(x_2(3) - x_2(2))(x_1(1) - x_1(2)) + (x_1(3) - x_1(2))(x_2(2) - x_2(1))}.$$

The weights $\alpha_1(i)$, $\alpha_2(i)$ and $\alpha_3(i)$, with $0 \le \alpha_1(i), \alpha_2(i), \alpha_3(i) \le 1$ and $\alpha_1(i) + \alpha_2(i) + \alpha_3(i) = 1$, are the contributions of the known observations (i.e., the vertexes of the triangle**)** to the new data point **x**(i) and determine its position within the enclosing simplex. Any property *y*(i) of the new observation is estimated as a weighted sum of the properties of the known points:

$$\hat{y}(i) = \alpha_1(i) y(\mathbf{x}_1) + \alpha_2(i) y(\mathbf{x}_2) + \alpha_3(i) y(\mathbf{x}_3), \tag{2}$$

where $y(\mathbf{x}_1)$, $y(\mathbf{x}_2)$ and $y(\mathbf{x}_3)$ are the values of the property for the vertexes. Once the weights are calculated, the estimation of any property is immediate. It is also straightforward to generalize Equation 1 and 2 to more than two dimensions.

The special case is for the estimation of a new observation that does not fall inside the convex hull defined by the known data points; hence, not even in any of the constructed simplexes. In this situation Equation 1 does not hold as the convexity constraints cannot be satisfied. Some of these observations are outliers (in a strict sense) but can also be good data located in region of the input space that was unknown when the initial calibration set was defined. It is worthwhile noticing that, in this sense, the main

limitation of the DTR method is its near-absolute lack of extrapolation ability. However, this limitation is not as dramatic as it may appear, because of the simplicity to update both the projection and the tessellation to account for the *outlying* data points

For the estimation of the properties of such observations, several approaches are reported in the literature (Fernández Pierna *et al*, 2002, Fernández Pierna *et al*, 2003, Jin *et al*., 2003a and Jin *et al*., 2003b) and can be adopted. In this applicative study, an estimate for a new observation situated outside the mesh is obtained by projecting it onto the closest simplex, as identified by its centroidal point (i.e., an artificial observation with a set of identical positive weights that sum up to one). The property is then estimated as equal to what calculated for the centroid of such a simplex.

3. Estimation of aromatics in LCO

A total of 91 LCO and Hydrotreated LCO (HDT LCO) samples were acquired for the present study. The HDT LCO samples were obtained at Sartec S.p.A. in a bench-scale pilot unit (Vinci Technologies) operating at various temperatures and pressures, by processing from different LCO feeds provided by the Saras Refinery (Sarroch, Italy). The unit mimics the most typical industrial operations and ensures the range of variation in the total aromatic content and in the distribution of the mono-, di- and tri+ aromatic classes expected in the full-scale case. The NIR spectra of the samples were recorded using a Varian Cary 500 Scan double-beam spectrophotometer in the wavelength range 1600-800 nm (*1 nm* resolution). The aromatic content (*w%*), was determined with the HPLC method EN-12916 using a modular Agilent 1100 Series system with refraction index detection. For the development of the multivariate models, the observations were divided in a calibration and validation set (respectively, 58 and 33 data points) chosen in order to contain samples that provide examples of all products' qualities and spanning the entire range of variation in the concentrations of the aromatics.

3.1. Results and discussion

Based on the data points in the calibration set, a dimensionality reduction with the Principal Components Analysis was performed, as a first step. As discussed in Section 2, the technique of choice uses only the input observations (the NIR spectra, Fig. 1a). After mean centering the observations, PCA was performed and the calibration spectra projected, Fig. 1b. The number of retained principal components is two. The selection was performed by inspection of the eigenvalues of the covariance matrix of the data; the two retained directions account for over *90%* of the total variation in the input space.

Upon projecting the input observations in the calibration set onto the first two principal directions, a Delaunay tessellation was performed, Fig. 1c. Each element of the mesh is a triangular simplex; the set of simplexes is enclosed in a bi-dimensional convex hull.

Subsequently, also the testing observations were mean-centered and projected onto the principal components space (Fig. 1d) where topological regression is performed. Again, it is worthwhile noticing that being the DTR model the same (the weights are calculated only once, from Eq. 1) for all the properties to be estimated, only a single regression model (Eq. 2) is needed; thus, minimizing the calibration and maintenance tasks.

The regression model was found by resolving for the coordinates of the testing observations and, then, calculating the respective properties from the known measurements. As for the testing observations that do not fall inside the convex hull determined by the calibration points (only 5, and all rather close to the boundaries of the convex hull, Fig 1d), the centroids method was used. The estimation results for the independent set of testing observations are depicted in the Fig. 2 and reported in Table 1 for the mono-, di- and tri+ aromatic content. In addition, we are reporting also the

results obtained when estimating the density of the samples. Density (*g cm-3*) was measured according to analytical method ASTM-4052.
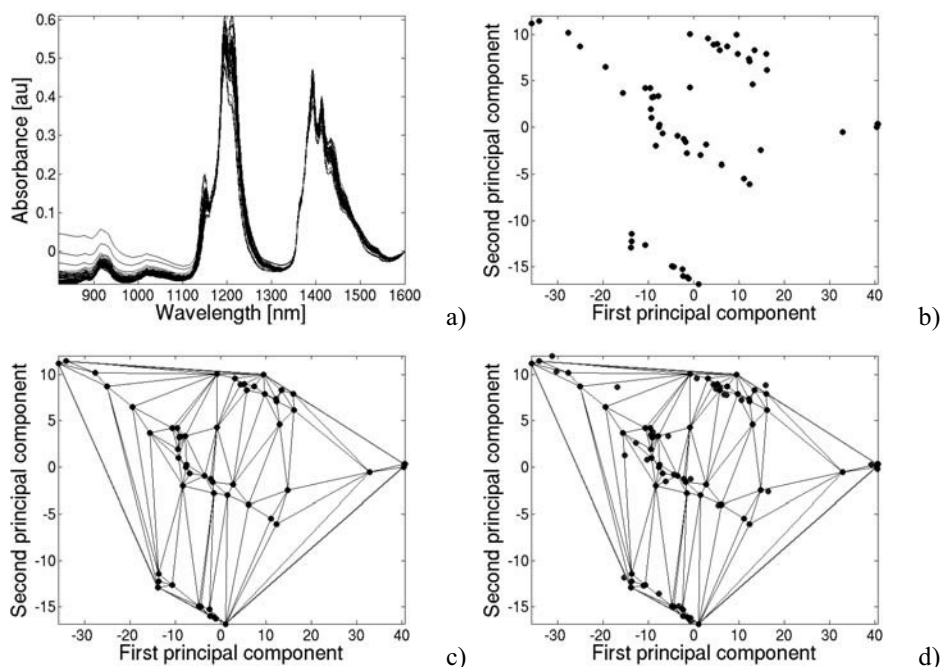
Figure 1. The spectral observations (calibration set in black and testing set in blue) in the original wavelength/absorbance space (a), projected in the principal components space (b), the mesh of simplexes generated by Delaunay triangulation (c) and, with the projected testing data points (d).

The accuracy of the estimates, expressed in terms of Root Mean Squared Error (RMSE) is within the repeatability range of the analytical measurements, and comparable with what obtained using a set of PLSR models independently cross-validated by Leave One Out (for the number of latent factors, 5). Table 1 also presents the results obtained with a set of PLSR models calibrated with 11 factors.

	Mono AH (*w%*)	**Di AH** (*w%*)	**Tri+ AH** (*w%*)	**Density** (*g cm^{-3}*)
DTR (2)	2.73	1.23	0.50	0.0021
PLSR (5)	2.06	0.98	0.80	0.0039
PLSR (11)	1.27	0.72	0.52	0.0021

Table 1. The estimation results (as RMSE) on the independent set of testing observations.

4. Concluding remarks

The results using the Delaunay Tessellation and Topological Regression presented here for the calibration of the aromatic content in Light Cycle Oils are at least comparable with the results of standard PLSR. The main advantages of the DTR method are in the simplicity of the calibration and ease to upgrade, but also the fewer components, thus leading to more robust and manageable models.
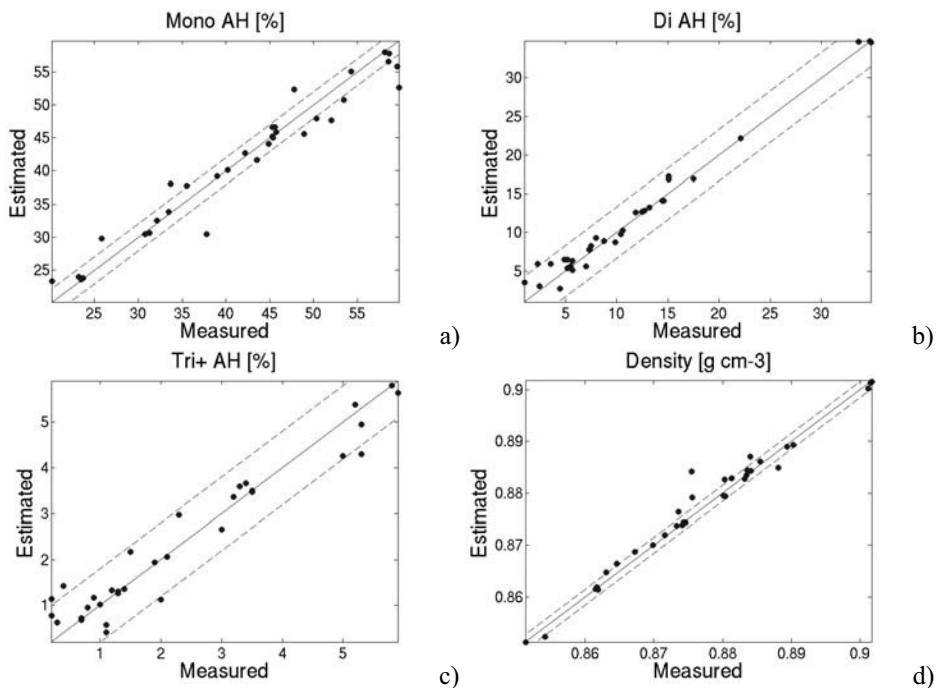
Figure 2. Estimation results (as scatter-plot) and the ASTM repeatability bands.

References

J.A. Fernández Pierna, L. Jin, M. Daszykowski, F. Wahl, D.L. Massart, 2003, A methodology to detect outliers/inliers in prediction with PLS, Chemometrics and Intelligent Laboratory Systems, 68, 17-28.

J.A. Fernández Pierna, F. Wahl, O.E. de Noord, D.L. Massart, 2002, Methods for outlier detection in prediction, Chemometrics and intelligent Laboratory Systems, 63, 27-39.

P. Gemperline et al., 2006, A Practical Guide to Chemometrics: Second Edition, CRC Press – Taylor & Francis.

J. Gudmundsson, M. Hammar, M. van Kreveld, 2002, High order Delaunay triangulations, Computational Geometry, 23, 85-98.

L. Jin, J.A. Fernández Pierna, F. Wahl, P. Dardenne, D.L. Massart, 2003, The Law of Mixtures method for multivariate calibration, Analytica Chimica Acta, 476, 73-84.

L. Jin, J.A., Fernández Pierna, Q. Xu, F. Wahl, O.E. de Noord, C.A. Saby, D.L. Massart, 2003, Delaunay triangulation method for multivariate calibration, Analytica Chimica Acta, 488, 1-14.

L. Jin, Q.S. Xu, J. Smeyers-Verbeke, D.L. Massart, 2005, Updating multivariate calibrations with the Delaunay triangulation method, Applied Spectroscopy, 59, 1125-1135.

L. Jin, Q.S. Xu, J. Smeyers-Verbeke, D.L. Massart, 2006, Updating multivariate calibration with the Delaunay triangulation method: The creation of a new local model, Chemometrics and Intelligent Laboratory Systems, 80, 87-98.

I.T. Jolliffe, 2002, Principal Components Analysis: Second Edition, Springer.

J.A. Lee, M. Verleysen, 2007, Nonlinear Dimensionality Reduction, Springer.

J.J. Jr. Workman, 1999, Review of process and non-invasive near-infrared and infrared spectroscopy: 1993-1999, Applied Spectroscopy Reviews, 34, 1, 1-89.

An Evolutionary Approach to Derive Adaptive Optimal Control Policy for Chemical Reaction Processes

Yoshiaki Shimizu

Department of Production Systems Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, tenpaku-cho,Toyohashi 441-8580, Japan

Abstract

To obtain a near optimal control policy for real world chemical processes, in this paper, we focused our attention on a new meta-heuristic method termed differential evolution (DE). Compared with conventional approaches characterized by variational logic as well as by the inconveniences of simultaneous optimization methods for differential-algebraic equation, we have shown DE's ability to derive a near optimal solution adaptive to various requirements in practice. Since the algorithm is straightforward and flexible to manage various conditions that other conventional approaches could not cope with effectively, we can cope with high dimensionality and outstanding non-linearity peculiar to chemical process in real world applications. In numerical experiments, we provided three popular reaction processes, applied the proposed method under various meaningful conditions, and validated its adaptability in comparison with other methods.
Keywords: Process system, Optimal control, Differential Evolution, Piece-wise constant, Chemical reaction process.

1. Introduction

By virtue of the advanced progress of meta-heuristic optimization method, we are now ready to derive a near optimal control policy for multi-dimensional and complex real world applications in chemical processes. With such understanding, in this paper, we focus our attention on a new meta-heuristic method termed differential evolution (DE). In comparison with the conventional variational approaches, it can derive a near optimal solution efficiently but very simply without any differential information and additional relations like adjoint equations in Maximum Principle (MP) of Pontryagin (1962). Since it is also free from the complicated usage in simultaneous optimization methods, we can cope with high dimensionality, discontinuity, and outstanding non-linearity peculiar to real world chemical reaction processes.

The aim of this paper is to facilitate a wide application of simulation-based optimization methods like DE for optimal control problems. Finally, its manifold effects are validated through numerical experiments.

2. Problem statements

Optimal control problems , which belong to a class of the off-line dynamic optimal control problem, has been solved traditionally by a certain variational method innovated by the MP. Since singular control and state inequality constraint problems often appear in practical operations of chemical processes, it is almost impossible to solve high dimensional problems for such cases.

Roughly speaking, numerical methods to cope with differential-algebraic equation (DAE) optimization problems are classified into sequential and simultaneous strategies

(Biegler, Cervantes and Wachter, 2002). The former is a control variable parameterization method while the latter uses the difference equation forms both of control variables and state variables. Therefore, the simultaneous approach needs to solve an extremely large-scale non-linear programming problem though it can handle the algebraic constraints by the same manner, i.e., difference equations.

On the other hand, in the sequential or control variable parameterization approach, the search

Fig.1 Piece-wise constant control

space is limited within a small region since the differential equations are not transformed into difference equations. However, it is not straightforward to handle the algebraic constraints.

Since, in many chemical processes, some controlling actions are operated discontinuously along time or space co-ordinate, this sequential approach should be highlighted by virtue of recent progress of optimization methods. Among them, piece-wise constant control (PWCC) (Fig.1) is one of the control classes that is easy to implement in reality. Hence, some meta-heuristic methods were successfully applied to obtain the optimal PWCC (Cruz et al., 2003; Kapadi & Gudi, 2004). However, these studies simply applied the methods to search only value of control variables and neglected an interval search and various practical conditions.

Since there has never existed any satisfactory answer to obtaining the optimal control from a practical point of view, it is essential to develop an approach realizing adaptive application to large-scale and complex systems in a flexible and straightforward manner.

2.1. Fixed-time Discrete Problem

Let us consider a system whose dynamics are described by

$$\dot{x} = f(x(t)), u(t), t), \quad x(t_0) = x_o \tag{1}$$

where $x(t)$ denotes an n-dimensional state vector, $u(t)$ an r-dimensional control vector, $f(\cdot)$ an n-dimensional vector function and t a time or time analog. Assuming a scalar control variable $u(t)$, for simplicity of description, the control policy of PWCC is prescribed by a set of pair (v_i, τ_i), where v_i is the value of $u(t)$ at i-th interval and τ_i is the i-th switching point. Then, the continuous problem can be transformed into the following 2N-1 dimensional optimization problem where v and τ are the vector forms of the sequences of v_i (i=1,2,···,N) and τ_i (i=1,2,···,N-1), respectively.

$$(\text{p.1}) \quad \underset{v,\tau}{Max} \quad J = K(x(\tau_N), \tau_N) \quad \text{subject to} \quad \begin{cases} \dot{x} = f(x(t), v_i), (\tau_{i-1} \leq t < \tau_i) \\ x(\tau_{i-1}^+) = x(\tau_{i-1}^-) \\ \underline{u}_i \leq v_i \leq \overline{u}_i \quad (i = 1, \ldots, N) \end{cases}$$

where $\underline{u}$ and $\overline{u}$ are lower and upper bound vectors of $u(t)$, and N denotes the number of division and controls the fineness of the control policy. In the case of equivalent interval, i.e., $\tau_j - \tau_{j-1} = \tau_{j+1} - \tau_j$, ($i$=1,2,···,$N$-1), the dimension of search problem (v-mode search) reduces to just N, apparently. Moreover, it should be noted that $\tau_0 = t_0$ and $\tau_N = t_f$ are implied, and each τ_i ranges from 1 to N_T, satisfying the condition that $\tau_j > \tau_{j-1}$, ($j = 1, \ldots, N$) where $N_T = (t_f - t_0)/H$ (H is an increment width of the numerical solution of the differential equation).
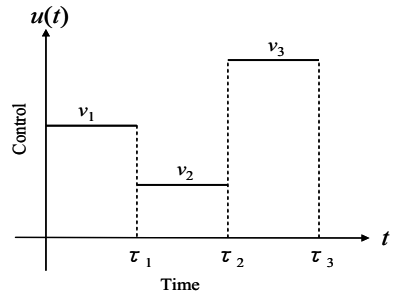
2.2. Free-time Discrete Problem

A typical example of the free-time problem is known as the time optimal control problem described in the following.

(p.2) Min $J=t_f$ subject to $\begin{cases} Eq.(1) \text{ and } \underline{u} < u(t) \leq \bar{u} \\ x(t_f) = x_f \end{cases}$

Here let us note the additional boundary condition of the state variables at the final time. The present simulation-based approach is almost impossible to always satisfy this condition directly. To cope with this problem, we use the penalty function method by augmenting the boundary condition into the objective function as follows.

(p.3) $\qquad t_f + P \left\| x(t_f) - x_f \right\| / \left\| x_f - x_0 \right\|$

where $\| \ \|$ denote a certain norm of vector, and P a penalty coefficient whose value should be increased adaptively according to the following formula.

$$P := \begin{cases} P + \Delta p & if \quad G^k < G^{k-1} & (coming\ up\ the\ target) \\ P & if \quad G^k \geqq G^{k-1} & (leaving\ the\ target) \end{cases} \qquad (2)$$

where G^k denotes the value of $\left\| x(t_f) - x_f \right\| / \left\| x_f - x_0 \right\|$ at the k-th iteration, and Δp a certain incremental amount.

3. Evolutionary Optimization to Derive Adjustable Optimal Control

3.1. Application under various control policies

For solving (p.1) and (p.2), it is straightforward to apply DE in the following manner. The genetic code of DE is made of the sequence of v_i $(i = 1, \ldots, N)$ for the equal interval problem, and in addition to it, the set of the interval length, i.e., $z_i = (\tau_i - \tau_{i-1}) / H$, $(i = 1, \ldots, N)$ for the unequal interval problem. Once these values are prescribed, we can evaluate the objective function directly after solving the differential equations numerically. Then, we can utilize simply this value as the fitness of the DE. Since the interval length z_i should be integer, rounding-off is carried out to satisfy this condition. The initial population is generated randomly by scattering the value between the lower and upper bounds of the decision variables.

Generally speaking, when N is small, the control often changes rapidly between the consecutive sequences. Hence, it likely happens that actual control action cannot catch up such a large change instantly. When this is the case, it is possible to impose the following condition.

$$|v_{i+1} - v_i| / T_{setup} \leq v_{speed}, (i = 1, \ldots, N-1) \qquad (3)$$

where T_{setup} and v_{speed} denote a transient interval and an admissible switching speed of the control policy, respectively. Similarly, when the duration interval of each v_i is prescribed, we can design the strategy by imposing conditions such that $\tau_j - \tau_{j-1} \geq D$, $(i=1,2,\ldots,N)$ for the unequal interval case and $(t_f - t_0)/N \geq D$, for the equal interval case. Here D denotes the minimum duration interval.

On the other hand, when N becomes large, this often causes the chattering action within the bounded control region in numerical solution. To prevent such chattering, the condition $|v_{i+1} - v_i| \leq \varepsilon$, $(i = 1, \ldots, N-1)$ is available to smooth the control policy, where ε is an appropriate small positive value.

3.2. State inequality constrained problem

Chemical processes are often controlled by manipulating flow rates and are restricted within a certain state space prescribed by temperature, pressure, and concentrations of chemical species, for examples. The latter condition requires us to augment the additional algebraic conditions like $g(x(t), u(t)) \leqq 0$, $t_0 \leq t \leq t_f$ into the original model. Instead of the tedious procedures discussed previously (Takamatsu and Shimizu, 1978a), the present approach can manage this problem

Fig.2 Control policies for various N
(Only v-mode search when N=50 and 100)

much more simply even for this DAE system. That is, it is enough to degrade the fitness value $Fit(v, \tau)$ of the DE according to the degree of violation of the state-inequality condition in the course of the transition as

$$Fit'(v, \tau) = Fit(v, \tau) + P \cdot \sum_{i=1}^{N} \delta_i(g)$$, where P denotes the penalty coefficient and the

psedo-heaviside function δ_i takes the value only when anyone of the state inequality constraints are violated at the i-th interval.

3.3. Insensitive control against parameter deviations

Describing explicitly a parameter vector p of the system model, let us note the existence of uncertainty in the model. Then we tried to obtain an insensitive control policy against the uncertain parameters. For this purpose, we derive the sensitivity equations with respect to p. Adjusting a certain norm of the sensitivity in an appropriate manner, e.g., $\|(\partial K / \partial x)(\partial x(t_f) / \partial p)\|$ for the objective function value, $\|(\partial x(t) / \partial p)\|$ for each state variable, and $\|(\partial g / \partial x)(\partial x(t) / \partial p)\|$ for the state-inequality constraints, for examples, we can give a suitable answer for the problem under consideration. Finally, this insensitivity goal is augmented to the original objective function J as $J' = wJ + (1 - w)S$, where S represents the sensitivity term given by the appropriate norm exemplified in the above, and w is the weighting factor to adjust the trade-off between the two goals, i.e., J and S.

4. Numerical Experiments

In the following numerical experiments, DE/rand/1/bin (Storn & Price, 1997) is applied. After some preliminary numerical experiments, we set the tuning parameters as F=0.8, P_c=0.5, N_p=3(2N-1) for the unequal interval problem while 3N for the equal one and the sopping generation=20N_p^2, where F, P_c and N_p denote the scale factor, the cross-over rate and the population size, respectively. Under these conditions, we observed a good profile of convergence all in the following numerical experiments.

4.1. Temperature control of completely mixed batch reactor (A)

For the reaction scheme like $A \xrightarrow{k_1} B \xrightarrow{k_2} C$, the optimal temperature control problem that maximizes the production of the desired product B is well known and is formulated as follows.
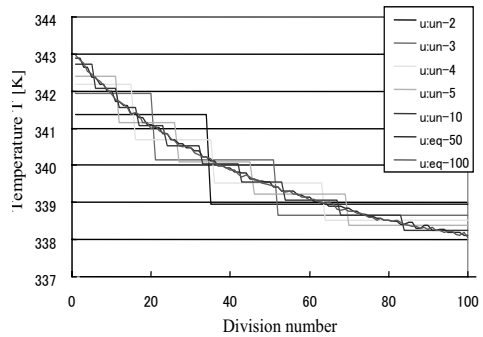
$$\underset{T}{Max} \quad J = x_2(t_f) \quad subject\ to \begin{cases} dx_1/dt = -k_1x_1, \qquad x_1(0) = x_{10} \\ dx_2/dt = k_1x_2 - k_2x_2,\ x_2(0) = x_{20} \end{cases}$$

where x_1 and x_2 are the concentrations of components A and B, respectively, t_f is the batch time, and k_1 and k_2 are the reaction rate constants of Arrhenius type.

As shown in Fig.2, the DE derived the solutions efficiently from $N=2$ up to the minimum grain interval problem, i.e., $N=N_f=100$. The only v-mode search has been done for $N=50$ and 100. When $N=100$, the result is almost equivalent to that of the variational method.

4.2. Catalyst blending problem (B)

The catalyst blending problem studied originally by Hofer (1976) is known to involve a singular control portion. With the prior knowledge that the optimal control is composed of three segment, i.e., upper bound, constant singular, and lower bound segments, a sensitivity method (Takamatsu and Y. Shimizu, 1978b) can derive the result very quickly by virtue of the differential information and small problem

Fig.3 Result of control policy

size. However, without such prior information, the DE outperformed this sensitivity method. Figure 3 depicts the results obtained by the proposed method. They are nearly identical to the profile of the sensitivity method when $N=3$ which is, in turn, almost same as the optimal control.

4.3. The flow rate control under state inequality constraint (C)

4.3.1. Fixed-time Solution

For the first-order reversible exothermic chemical reaction (Siebenthal and Aris, 1964), let us suppose that the temperature in completely mixed batch reactor with cooling jacket must not exceed the upper limit $\overline{T}$. The optimal cooling control problem that will achieve the maximum production of B is described with dimensionless variables as follows (Takamatsu and Shimizu, 1978a).

$$Max \quad J = x_1(1) \qquad subject\ to$$

$$\begin{cases} \dot{x}_1 = p_1(1-x_1)\exp(p_2/x_2) - p_3x_1\exp(p_4/x_2), \quad x_1(0) = x_{10} \\ \dot{x}_2 = p_5[p_1(1-x_1)\exp(p_2/x_2) - p_3x_1\exp(p_4/x_2)] - p_6u,\ x_2(0) = x_{20} \\ \underline{u} \le u(\tau) \le \overline{u}, \quad x_2(\tau) \le \overline{x}_2 \quad (0 \le \tau \le 1) \end{cases}$$

where the kinetic rate constants are of the Arrhenius type. Due to the state-inequality constraint over the time horizon and the linearity of the control variable, this problem is known to be very difficult to solve.

Numerical results revealed that the good performance of the simplex method of Nelder and Mead is limited only to the small N, and the unequal interval search with moderate $N (= 5)$ is enough to attain a satisfactory performance level (the gap to the optimal value is less than 0.3%). The chattering becomes heavily for N greater than 50 to imitate the optimal trajectory.

4.3.2. Insensitive control against parameter deviations

It is likely that the fouling degrades the performance of the cooling jacket, and the reactor temperature may exceed the limit. To prevent such defect a priori, an insensitive control policy against the parameter deviations (fouling) is more desirable to keep the safety or the long life of the reactor. To realize this aspect, the sensitivity

equations with respect to p_6 are derived. The present strategy for the insensitive control is to take an action when the state trajectory approaches the limit beyond a certain threshold value $\Delta x_2 (= 0.01\,\overline{T}/T_0)$. That is, $\|(\partial g/\partial x)(\partial x(t)/\partial p)\|$ in section 3.3 is defined as $\left\{\int_{t_2}^{t_1} (\partial x_2(t)/\partial p_6)^2\,dt\right\}^{1/2}$ where t_1 and t_2 prescribe the beginning and the ending points of the interval featuring the situation, $\bar{x}_2 - x_2(t) \le \Delta x_2$. By considering the insensitivity, the state trajectory tends to keep inside of the upper limit of temperature as illustrated Fig.4. Accordingly, the violation does not occur even with the 20% parameter variation ($p_6' = 0.8\,p_6$). Moreover, it only causes a small degradation of the system performance both at the nominal (ordinary=0.9153 vs. insensitive=0.9143) and at the deviated (0.9074 (infeasible) vs. 0.9065).

4.3.3. Free Time Solution

The objective function of this problem will be changed as $Min \quad \tau_n \quad (\tau_n \le \tau_N = 100)$. Additionally, we need to add the terminal condition like $x(\tau_n) = x_f$. For three problems with different terminal conditions shown in Table 1, the proposed approach solved all cases within the square root error 0.0001 of the terminal condition. Figure 5 depicts the control policy for Run 1 when $N=10$.

Table 1 Numerical results for free time problem

Run	x_f	τ_n [-]
1	(0.88, 0.92)	63
2	(0.9, 0.9)	83
3	(0.9, 0.95)	86

($N=10$; unequal interval policy)
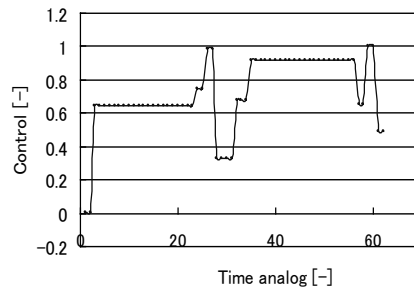
Fig.5 Profile of control for Run 1

5. Conclusion

To derive the optimize control policy for multi-dimensional and complex systems aimed at real world applications, we focus our attention on a new meta-heuristic method termed DE. Compared with the conventional approaches, we have shown DE's ability to derive optimal PWCC efficiently under various conditions popular in real world applications. In comparison with other methods through various numerical experiments, we have confirmed the great promise of DE as a practical solution method for optimal control problems.

References

L. T. Biegler, A. M. Cervantes and A. Wachter, 2002, Chem. Eng. Sci., 57, 575-593

I.L. L.Cruz, , L.G. Willigenburg and G. Straten, 2003, Applied Soft Computing, **3**, 97-122

M. D. Kapadi and R. D. Gudi, 2004, Process Biochemistry, **39**, 1709-1721

R. Storn and K. Price, 1997, Journal of Global Optimization, 11, 341-359

T. Takamatsu, Y. Shimizu and T. Murata, 1978a, J. of Chem. Eng. of Japan, 11, 59-66

T. Takamatsu and Y. Shimizu, 1978b, J. of Chem. Eng. of Japan, 11, 221-226
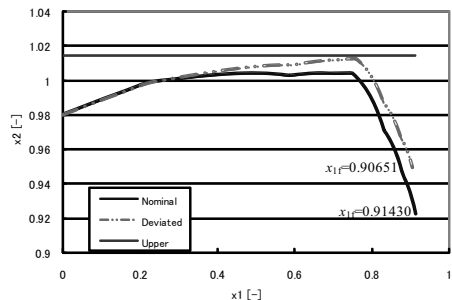
Fig.4 Nominal and deviated trajectories

A Novel Technique to Estimate Valve Stiction Based on Pattern Recognition

"Marcelo Farenzena, Jorge O. Trierweiler"

"Department of Chemical Engineering, Federal University of Rio Grande do Sul,"
Porto Alegre, RS, 90040-040, Brazil

Abstract

This work proposes a novel method to estimate valve stiction parameters that requires only normal operating data (process variable (PV) and controller output (OP)). The technique is based on pattern recognition, where a neural network, called Stiction Inference Model (SIM), predicts the stiction parameters. Thus, the quantification is very fast and robust, what allows the application in the industrial scenario, where a large number of loops should be analyzed. The SIM is described, as well as the procedure to its construction: the dataset construction, the variable selection, and the network training. The SIM obtained has a compact size (10 neurons) and a good performance for both parameters ($R^2=0.99$). The proposed method was applied in a set of valves, providing reliable results.

Keywords: Process control; Process monitoring; Valve stiction; Hysteresis.

1. Introduction

The diagnostics of the "valve health" is important to eliminate plant oscillations, allowing the process to achieve a more profitable operating point. The villain in chemical plants is high static friction (or stiction). Reports show that 20-30% of all control loops have bad performance because of stiction (Bialkowski, 1993). That is why preventive maintenance is essential to avoid off-spec products and waste of energy or raw materials. This scenario demands an automatic technique to detect and quantify valve.

In the last ten years, many works and methods have been proposed to diagnose and solve this valve disorder. A first group of works aimed at diagnosing stiction automatically, using only process variable (PV) and controller output (OP) (He et al., 2007; Horch, 1999; Rossi and Scali, 2005; Singhal and Salsbury, 2005; Yamashita, 2006). Some works have proposed specific stiction models for diaphragm type valves. A good survey about stiction models was recently written by Garcia (2008).

Furthermore, some authors have proposed to compute stiction parameters in real time. Choudhury et al. (2006) proposed two methods to quantify stiction parameters, based on ellipse fitting and *c-clustering*, using a one parameter empirical model. Subsequently, some authors have proposed to quantify stiction parameters using a more reliable model, with two parameters. Choudhury *et al.* (2008) have proposed a method based on optimization and grid search. Jelali (2008) has introduced one methodology to compute both stiction parameters, based on least-squares and global search algorithms. This method is computationally expensive what restrict its on-line application.

The main contribution of this work is to overcome this limitation, proposing a method to compute the two stiction parameters, where the stiction parameters are estimated with a very fast computational time. The technique is based on pattern recognition, where a neural network, called Stiction Inference Model (SIM), predicts the stiction parameters.

Thus, the quantification is very fast and robust. Moreover, the method needs only normal operating data, i.e. based on process variable (PV) and controller output (OP) patterns. No information about the valve stem is necessary. Based on these premises, the method can be applied industrially, where a large number of valves is analyzed. The procedure to build the SIM has used the stepwise regression to select the input variables and a neural network to capture the desired behavior input-output The SIM accuracy is slightly poorer – with less than 30% relative error when compared to the least-squares and global search algorithm. However, our objective is to propose a fast algorithm to allow its large-scale application.

The paper is segmented as follows: in section 2 the stiction phenomenon is defined and modeled. In the next section, Stiction Inference Model (SIM) is defined, as well as the procedure to build it. Next, in the section 4, the SIM is applied in a set of case studies. The paper ends with the concluding remarks.

2. Stiction: definition and model

The phenomenon of stiction can be summarized as follows (Ruel, 2000): "*Stiction is the resistance to the start of motion, usually measured as the difference between the driving values required to overcome static friction upscale and down scale. The word stiction is a combination of the words stick and friction, created to emphasize the difference between static and dynamic friction. Stiction exists when the static (starting) friction exceeds the dynamic (moving) friction inside the valve. Stiction describes the valve's stem (or shaft) sticking when small changes are attempted. Friction of a moving object is less than when it is stationary. Stiction can keep the stem from moving for small control input changes, and then the stem moves when there is enough force to free it. The result of stiction is that the force required to get the stem to move is more than is required to go to the desired stem position. In presence of stiction, the movement is jumpy.*"

2.1. Stiction model

A sticky valve has in the phase plot (Valve output (MV) versus Controller output (OP)), shown in Fig. 1, four components: deadband (*DB*), stickband (*SB*), slip jump (*J*) and moving phase (*MP*). The method assumes that the process and controller have linear behavior, while the sticky valve inserts in the loop nonlinear behavior.

When the valve changes the direction (A), the valve becomes sticky. The controller should overcome the deadband (AB) plus stickband (BC), and then the valve jumps to a new position (D). Next, the valve starts moving, until its direction changes again or the valve comes to rest, between D and E.

The stiction model consists of these two parameters: S, called staticband, (deadband+stickband) and J (slipJump). The deadband and stickband represent the behavior of the valve when it is not moving, although the input of the valve keeps changing. The magnitude of staticband and slipjump is essential to determine the limit cycle amplitude and frequency.
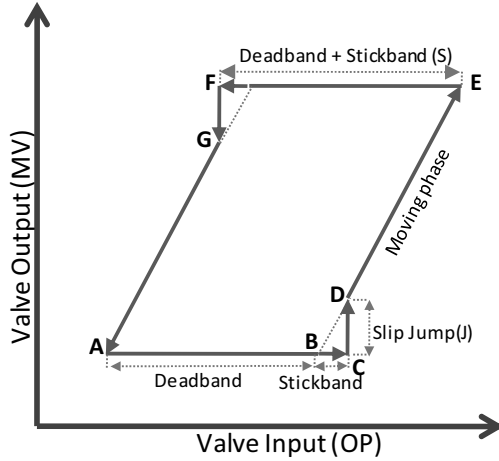
Fig. 1. Relation between controller output (OP) and valve position (MV) for a sticky valve.

The stiction model used in this work is proposed by Kano (2004), which is an extension of Choudhury's method, where stiction is modeled using two parameters. Their main advantage is that it can deal with both stochastic and deterministic signals.

2.2. Stiction evaluation

In the literature, two methods are available to compute the stiction parameters, using a two parameter model. The method proposed by Jelali (2008), a two step procedure is introduced. In the first step the stiction parameters are quantified using pattern search methods or genetic algorithms (GA). Following, the low-order linear plant model is identified, using a least-squares estimator. A second work of Choudhury *et al.* (2008) proposes a method based on a grid search. Initially, a grid using several different values of J and S is built and then based on the process output, the plant model is identified. Based on the mean square error (MSE) between predicted process output and actual output in each grid point, the stiction parameters are estimated.

3. Stiction Inferential Model (SIM)

The Stiction Inference Model (SIM), proposed in this work, can be defined as an inference model to evaluate the Staticband (S) and the SlipJump (J), whose inputs are parameters obtained only from normal operating data.

The procedure to construct SIM is analogous to the procedure to build a softsensor using neural networks. The steps are following listed and they will be explained in the next sections.

1. Dataset design: select plant models, disturbances, controller performance, and stiction parameters;
2. Generate the dataset and compute candidate inputs and outputs parameters;
3. Divide the dataset into test and training datasets;
4. Select the input variables for each stiction parameter;
5. Train the neural networks using, with variable number of neurons in the hidden layer;
6. Evaluate network performance for both datasets;

Each step will be described in details in the following subsections.

3.1. Dataset design

The dataset was designed based on a SISO loop, whose scheme is shown in Fig. 2.
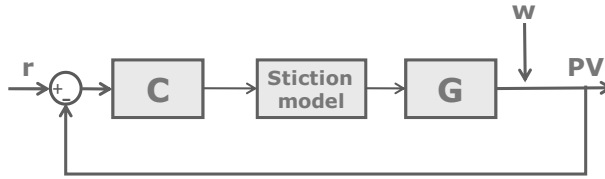
Fig. 2: SISO loop scheme

Where C is the PI controller, G is the first order plus time delay plant. The input w is white noise and r is the reference. In this work, we propose only a stiction inference model for first order plus time delay models. However it could be extended to higher order models or integrators. Tab 1 shows the variable parameters to generate the dataset and the interval for each parameter.

Tab. 1: Variable parameters used in the SIM data generation

Param.	Description	Interval
K_p	Controller gain	[0.1: 0.2: 1.1, 2: 1: 5]
τ_I	Controller integral constant	[0.3τ: 0.3τ: 3τ]
τ	Process time constant	[5: 5: 20, 30: 10: 100]
S	Static band	[0.5: 0.5: 5]
J	Slip Jump	[0.5: 0.5: 5]
$\sigma^2(w)$	White noise variance	[100 10 1]

The number of different scenarios in the dataset is 360,000. Subsequently, the complete dataset was randomly divided in two datasets - training dataset = 80% of the points and test dataset = 20% of the points

3.2. Input variables

The candidate variables to build the inference model and the procedure to select the best subset of them to describe the stiction are subsequently listed:

1. Δ_{PV}: Difference between the maximum and minimum value in PV;
2. Δ_{OP}: Difference between the maximum and minimum value in OP;
3. Ratio between Δ_{PV} and Δ_{OP}.
4. E_W: integral of square error of difference between PV and the best wave-shape curve interpolation (PV_W) (see (Rossi and Scali, 2005));
5. E_T: integral of square error of difference between PV and the best triangular curve interpolation (PV_T);
6. ZC: Number of zero-crossings in the zero-mean data;
7. ZC_{ACV}: Number of zero-crossings in the autocovariance function (Thornhill et al., 2003).

The variables were selected using stepwise regression (Rawlings, 1988). The inputs selected to estimate J are the same as the inputs to predict S:

- Δ_{PV}
- Ratio between Δ_{PV} and Δ_{OP}
- ZC_{ACV}

3.3. Neural-network training

To interpolate the training dataset, we use a feedforward backpropagation neural network (Haykin, 1999), from Matlab 5.3 (R11, Neural Network toolbox ver. 3.0.1). In the hidden layer, hyperbolic tangent sigmoid neurons were used, while in the output layer, linear neurons are used. In the training step, the layer performance is measured using mean square error. The training function used was Levenberg-Marquardt backpropagation (*trainlm*) and the learning function was Gradient descent with momentum weight and bias (*learngdm*).

We try networks with different number of neurons in the hidden layer to select the one that better predict the stiction parameters, with the minimum number of neurons. Initially, the SIM for Staticband (S) was designed. Tab. 2 shows the networks performance, with different number of neurons, for the training and test datasets.

Tab. 2: networks performance – output variable: Staticband (S)

Neurons	R^2 Training dataset	R^2 Test dataset
5	0.99	0.99
10	0.99	0.99
20	0.99	0.99
50	0.99	0.99

Based on Tab. 2, the best net configuration has 5 neurons in the hidden layer. Subsequently, the model for Slipjump (*J*) was designed. Tab. 3 shows the networks performance, with different number of neurons, for the training and test datasets.

Tab. 3: networks performance – output variable: Slipjump (J)

Neurons	R^2 Training dataset	R^2 Test dataset
5	0.97	0.97
10	0.98	0.98
20	0.98	0.98
50	0.98	0.98

Again, the best configuration, according to Tab 3, has 10 neurons in the hidden layer.
The final SIM structure for both parameters is:

- Neural network model with 2 layers;
- 1st layer (hidden): 10 hyperbolic tangent sigmoid neurons;
- 2nd layer (external): 1 linear neuron.

4. Case studies

This section illustrates the SIM application using simulation datasets. The comparison between the predicted and real values for stiction parameters will be made for a set of 6 valves, which are not used to train or test the SIM neural networks. Tab. 4 shows the results for original *J* and *S*, and its predicted values (J_P and S_P) by the SIM.

Tab. 4 clear shows the good performance of SIM, where the relative error for all controllers was lower than 30%. Using a Pentium D, 1 GB ram, the computation time for all cases was lower than 1 second. Using the same computer, to predict valve

stiction to 100,000 valves took 1 second, what allows the real-time and large-scale application of the proposed methodology.

Tab. 4: SIM prediction for 6 different sticky valves

Cont.	S	S_p	J	J_p
Cont1	10	13	9	9,9
Cont2	20	23	12	11
Cont3	20	20	15	15
Cont4	20	27	40	39
Cont5	30	33	30	30
Cont6	40	37	60	53

5. Conclusions

In this work, a novel algorithm to evaluate valve stiction, called Stiction Inference Model (SIM), has been proposed. It allows the computation of both stiction parameters (staticband S and slipjump J) using only routine operating data. The SIM is based on neural networks. Comparing the proposed method with to the least-squares and global search algorithm, the SIM accuracy is slightly poorer. However, its large-scale application is viable, because of its computational time.

References

W.L. Bialkowski,1993, Dreams versus reality: A view from both sides of the gap, Pulp and Paper Canada, 94, 11, 19-27.

M.A.A.S. Choudhury, M. Jain, and S.L. Shah, 2006, Detection and quantification of valve stiction, Proceedings of the American Control Conference, Vol. 2006, pp. 2097-2106.

M.A.A.S. Choudhury, M. Jain, and S.L. Shah,2008, Stiction - definition, modelling, detection and quantification, Journal of Process Control, 18, 3-4, 232-243.

C. Garcia,2008, Comparison of friction models applied to a control valve, Control Engineering Practice, 16, 10, 1231-1243.

S.S. Haykin, 1999, Neural networks : a comprehensive foundation, Prentice Hall, Upper Saddle River, N.J.

Q.P. He, J. Wang, M. Pottmann, and S.J. Qin,2007, A curve fitting method for detecting valve stiction in oscillating control loops, Industrial and Engineering Chemistry Research, 46, 13, 4549-4560.

A. Horch,1999, A simple method for detection of stiction in control valves, Control Engineering Practice, 7, 10, 1221-1231.

M. Jelali,2008, Estimation of valve stiction in control loops using separable least-squares and global search algorithms, Journal of Process Control, 18, 7-8, 632-642.

M. Kano, H. Maruta, H. Kugemoto, and K. Shimizu, 2004, Practical model and detection algorithm for valve stiction, 7th IFAC DYCOPS,IFAC, Boston, USA.

J.O. Rawlings, 1988, Applied regression analysis : a research tool, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, Calif.

M. Rossi, and C. Scali,2005, A comparison of techniques for automatic detection of stiction: simulation and application to industrial data, Journal of Process Control, 15, 5, 505-514.

M. Ruel,2000, Stiction: The hidden menace, Control Magazine.

A. Singhal, and T.I. Salsbury,2005, A simple method for detecting valve stiction in oscillating control loops, Journal of Process Control, 15, 4, 371-382.

N.F. Thornhill, B. Huang, and H. Zhang,2003, Detection of multiple oscillations in control loops, Journal of Process Control, 13, 1, 91-100.

Y. Yamashita,2006, An automatic method for detection of valve stiction in process control loops, Control Engineering Practice, 14, 5, 503-510.

Two-Layer Planning and Scheduling of Batch Production Processes Under Uncertainty

Martin Hüfner, Thomas Tometzki, Sebastian Engell

Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, TU Dortmund, 44221 Dortmund, Germany

Abstract

This paper proposes a new combined algorithm to solve planning and scheduling problems of batch production processes under uncertainty. For the planning problem under uncertainty, a two-stage stochastic integer program is formulated and solved by a stage-decomposition algorithm. To evaluate the feasibility of a production plan, a priced timed automata model is set up for the scheduling problem of the next planning period and solved using reachability analysis. If the targets from the planning layer cannot be met, a penalty term is returned to the planning layer which is used to re-evaluate the candidate production plans. The two-layer concept is applied to a multi-product batch plant demonstrator.

Keywords: evolutionary algorithms, MILP, timed automata, multi-product batch plants

1. Introduction

For the economic operation of multi-product batch plants, high-quality *production plans* that reflect the uncertainties in market and technical parameters as well as feasible or optimal or at least feasible *operation schedules* are needed. *Production planning* is understood as the decision *which* products have to be produced in the planning periods with the limited set of plant equipment. *Scheduling* concerns the decisions *how* the scarce resources are allocated to the operations over time to produce the planned products in order to meet the targets. The combination of long-term production planning and of scheduling ensures the meeting of the economic objectives and timely completion of the tasks associated with the production plans. On the other hand, it is not necessary to determine precise schedules over long horizons because the uncertainties render such schedules obsolete after short periods of time.

The decision structure in production planning with uncertainties is reflected well by a mixed-integer recourse model with a finite number of scenarios in the form of a *two-stage stochastic integer program* (2-SIP) [1]. The recourse property reflects the ability to compensate the effects of a set of *here-and-now* (first-stage) decisions which have to be made under uncertainty by a second set of *recourse* (second-stage) decisions which can be made after the uncertainty has realized. This type of models is particularly appropriate for moving horizon production planning where the movement of the horizon is accompanied by the realization of uncertainties and their compensation.

In [2], a similar temporal decomposition approach to the coupled production planning and scheduling problem is applied, where first the time period and the products that are considered in short-term scheduling are determined, and then the scheduling problem is solved for the immediate future. In [3], a simultaneous solution of production planning and scheduling problem is addressed through a hierarchical framework. The planning problem aggregates orders in the planning period and considers uncer-

tainty utilizing multi-stage stochastic programming. Using a moving horizon strategy, the production targets for the current stage are provided to the scheduling layer, which is solved using a continuous-time MILP formulation.

Optimal scheduling of multi-product batch plants (resource allocation, sequencing and timing) has been in the focus of research for many years now. A promising new approach to scheduling problems is to model them by *timed automata* (TA) and to perform optimization by *reachability analysis* [4, 5]. Modeling is completely graphical and the problem can be represented in a modular and transparent fashion with independent small models for recipes, resources, etc. that are connected by synchronization labels and shared variables and can be composed automatically.

2. Two-layer algorithm for production planning and scheduling

In this contribution, the coupled problem of production planning and plant scheduling is solved by a two-layer algorithm, where the planning (upper-layer) problem under uncertainty is represented as a *two-stage stochastic integer program* (2-SIP) and the scheduling (lower layer) problem is solved using a *timed automata* (TA) approach. The algorithm for the upper layer problem (see Section 4) interacts with the algorithm for the lower layer problem (see Section 5) in a feedback loop.

The upper layer algorithm iteratively generates a set of production plans. After every iteration, the upper layer passes the first stage solution with the best objective value to the lower layer. The lower layer algorithm computes an optimal or at least a feasible schedule for the first-stage periods and returns a penalty value to the upper layer if the due dates are not met. The penalty value is added to the planning objective value of the passed solution. Thus another production plan from the set may have a better objective. Then the scheduling is repeated for the new best solution and the resulting penalty is passed to the upper layer. The procedure iterates until the best solution from the set has been found.

3. Multi-product batch plant case study

In order to illustrate the approach, a small demonstrator batch plant [6] is considered (see Fig. 1). The end-products blue (**B)** and green (**G)** are produced from three raw materials, yellow (**Y)**, red (**R)** and white (**W)**. Two batches of materials **Y** and **W** react to produce one batch of product **B**, similarly two batches of **R** and **W** react to produce one batch of product **G**. The plant consists of 3 stages. The first stage consists of three buffer tanks which are used to store the raw materials **Y, R** and **W**. The second stage consists of three reactors that perform the reaction process to produce the end-products. Each reactor can be filled from each raw material buffer tank in the first stage; and it is possible to produce either product **B** or product **G** in each reactor. After processing the materials, a reactor contains one batch of product. The third stage consists of two buffer

Fig. 1: Demonstrator batch plant

tanks which are used to store the end-products **B** and **G** exclusively. Once started operations must be finished without any interruption (non-preemptive scheduling). The production of one batch of a recipe (**B** or **G**) consists of 6 operations and involves timing

constraints between individual operations. After the materials are processed by the reactors, the end-products must be drained into the buffer tanks in the third stage imme-diately, imposing a zero-wait constraint between the operations.

The planning problem is based on a simplified discrete time multi-period model which represents capacity constraints and material balances. The buffer tanks in the first stage contain at most two batches of the raw materials. Each of the buffer tanks has a maximum capacity to store three batches of the respective products. Raw materials can be stored temporarily in a depot. The planning problem is subject to uncertainties in the product demands, revenues, production costs, and possible breakdowns of the reactors. The planning decisions which have to be made are the optimal choices of the purchases of raw materials and the batch numbers of products. The profit that has to be maximized is calculated from sales revenues, production costs, storage costs, and penalties for lateness. The scheduling problem is based on a continuous-time formulation containing all important plant characteristics such as non-intermediate storage policies, exclusive assignment of resources to tasks, intermediate release dates for the raw materials and intermediate due dates for the end-products. The purchased raw materials are released at time points which correspond to the beginning of the discrete periods. For each batch of a product a due date at the end of a period has to be met. The objective of the scheduling problem is to schedule the production orders with no delay.

4. Upper layer algorithm: Stage decomposition based hybrid approach

4.1. Two-stage mixed-integer program

The production planning problem under uncertainty is modeled as a 2-SIP. The uncer-tainty in the planning problem has $\Omega=32$ realizations that are modeled by a discrete set of scenarios $\omega = 1,\ldots, \Omega$. In a 2-SIP, the decisions are divided into the first-stage deci-sions $\mathbf{x}$ which have to be taken before the uncertainty is disclosed and second-stage decisions $\mathbf{y}_\omega$, which have to be taken after the uncertainty is realized.

$$\min_{\mathbf{x},\mathbf{y}_1,\ldots,\mathbf{y}_\Omega} \mathbf{c}^\mathrm{T}\mathbf{x} + \sum_{\omega=1}^{\Omega} \pi_\omega \mathbf{q}_\omega^\mathrm{T}\mathbf{y}_\omega \tag{1}$$

$$\text{s.t.} \qquad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{2}$$

$$\mathbf{W}_\omega \mathbf{y}_\omega \leq \mathbf{h}_\omega - \mathbf{T}_\omega \mathbf{x} \tag{3}$$

$$\mathbf{x} \in X, \mathbf{y}_\omega \in Y, \forall \omega=1,\ldots,\Omega.$$

The objective of the problem (1) consists of the first-stage costs and the expected value of the second stage costs. The costs are calculated as linear functions of the first-stage variables $\mathbf{x}$ and the second-stage variables $\mathbf{y}_\omega$ with vectors of cost parameters $\mathbf{c}$ and $\mathbf{q}_\omega$. The two-stage model consists of inequality constraints on the first-stage decisions (2) and on the first- and second-stage decisions (3). The finite sets X and Y may contain integrality requirements.

4.2. Stage decomposition approach

The main idea of stage decomposition [7] is to remove the ties between the second-stage scenario subproblems by fixing the first-stage decisions $\mathbf{x}$. The scenario subproblems are of significantly smaller size than the full two-stage problem. The master problem is a function of the vector of first-stage variables $\mathbf{x}$ only:

$$\min_{\mathbf{x}} \mathrm{f}(\mathbf{x}) = \mathbf{c}^\mathrm{T}\mathbf{x} + \sum_{\omega=1}^{\Omega} \pi_\omega Q_\omega(\mathbf{x}) \qquad \text{s.t.} \qquad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in X \tag{4}$$

The evaluation of the second-stage value function $Q_\omega(\mathbf{x})$ for a given $\mathbf{x}$ requires the solution of Ω independent MILP subproblems over the second-stage variables y_ω:

$$Q_\omega(\mathbf{x}) = \min_{y_\omega} \ \mathbf{q}_\omega^T \mathbf{y}_\omega \quad \text{s.t.} \quad \mathbf{W}_\omega \mathbf{y}_\omega \leq \mathbf{h}_\omega - \mathbf{T}_\omega \mathbf{x}, \ \mathbf{y}_\omega \in Y, \ \forall \omega = 1, \dots, \Omega. \tag{5}$$

The constraints of the master problem (2) are scenario independent, while the parameters of the second-stage constraints in (3) may vary from scenario to scenario. The vector of the first-stage variables $\mathbf{x}$ appears as a vector of fixed parameters in the constraints of the second-stage scenario problems. The main algorithmic idea is to address the master problem given by (4) by an *evolutionary algorithm*. To evaluate f($\mathbf{x}$), the Ω subproblems given by (5) are solved independently by a *MILP solver*.

4.3. Evolutionary Algorithm

The master problem (4) is tackled by an adapted (integer) *(μ, κ, λ)-evolutionary algorithm*. Each individual $\mathbf{x} = (x_1, \dots, x_n)$ of the population represents the choice of the numbers of batches of raw materials and batches of products for all first-stage periods. A population of μ individuals is initialized randomly within the bounds of the box-constrained first-stage feasible decision space $\mathbf{x}^{min} \leq \mathbf{x} \leq \mathbf{x}^{max}$. The fitness values of the initial population are calculated by the objective function f($\mathbf{x}$) of the master problem.

After the evaluation of the initial population, λ offspring individuals are generated by λ-fold application of the discrete recombination operator. The recombination chooses two individuals out of the μ parents with equal probability. Each x_i of the offspring is the corresponding value from the first or from the second parent with equal probability. After the recombination, the mutation operator mutates the λ offspring in two steps:

In the first step, the object parameters for the amounts of products are mutated. The reason for this is that the amount of products cannot exceed the amount of available raw materials. Each object parameter for the amount of a product x_i is perturbed independently from the others by a random number drawn from a normal distribution with an expected value of zero and a variance of one. This choice leads to an expected perturbation of 1 for each object parameter. For integer parameters, the random number is rounded to the nearest integer value. To maintain the bounds for the number of products, the search space is transformed into a periodic space, such that $x_i^{max} + 1 = x_i^{min}$.

In the second step, based on the amounts of products, the amounts of raw materials are generated. The minimum amount of raw materials which is needed to produce the desired amount of products is calculated for each period. Additional raw materials are added using the absolute value of random numbers from a discretized normal distribution. Beginning with the last period of the first-stage, the raw materials are shifted to earlier periods with low probability. In order to maintain feasible production plans, the amount of raw materials can only be shifted to earlier periods. All raw materials which are not needed can be stored.

The λ offspring candidate solutions are evaluated by the objective function f($\mathbf{x}$) of the master problem. The best solution $\mathbf{x}^*$ is passed to the *lower layer scheduling algorithm* (see Section 5). The lower layer algorithm computes an optimal schedule for the immediate future (first-stage periods) and returns the penalty value $0 < p(\mathbf{x}^*) < p_{max}$ if the due dates are not met or a large penalty value $p(\mathbf{x}^*) = p_{max}$ if no feasible schedule exists. The penalty value $p(\mathbf{x}^*)$ is added to the fitness value f($\mathbf{x}^*$). By adding the penalty to the objective value f($\mathbf{x}^*$) := f($\mathbf{x}^*$) + p($\mathbf{x}^*$), the fitness based order of the individuals in the current population may change. Then the new best solution is passed to the scheduling algorithm and a new optimal schedule is generated and the penalty costs are calculated. The procedure iterates until the first solution in the sorted population no longer changes.

In the selection for population replacement, a truncation selection chooses the μ best ($1 \leq \mu \leq \lambda$) individuals out of the union of μ parents and λ offspring which do not exceed the maximum age of κ for the next iteration. The age of an individual is increased by one after each generation. After the new population is created, a new iteration loop starts if the termination criterion is not fulfilled.

5. Lower layer algorithm: Timed automata based approach

5.1. Background of timed automata

The underlying scheduling problem is solved using a priced timed automata approach. Timed automata (TA) are *finite state automata* extended by the notion of *clocks* to model discrete event systems with timed behaviour. An extension of TA with the notion of costs is known as priced TA. A priced TA is a timed automata equipped with an additional function which assigns costs for staying in locations and costs for transitions. The resources, jobs and timing constraints are modelled as sets of priced TA. For the problem considered in the case study, release dates and due dates for the production orders exist. The release date for a production order is modelled by the guards on the transition that represents starting of the first operation. The due date for each production order is modelled by a separate TA (due date automaton) and synchronized with the respective job automaton. The delay cost for the production order which miss the due date are computed using the clock in the due date automaton. A detailed explanation of the approach to model release dates and due dates using TA can be found in [8].

5.2. Cost-optimal reachability analysis

Once the individual automata have been created, the *composed priced TA* is realized by the tool. The composed priced TA is automatically created by synchronizing the transition labels in the sets of the individual priced TA. TAOpt performs a symbolic reachability analysis to explore the solution space and to derive production schedules with minimal cost. The reachability tree is created from the composed automaton. It consists of nodes and transitions; a node represents a combination of a state and clock valuations of all clocks including the global clock. A cost-optimal reachability analysis is performed starting from the initial location where no jobs are started and trying to find a path to the target location where all jobs are finished within the defined due date. A depth first search algorithm is used to explore the reachability graph. The search space reduction techniques weak non-laziness, sleep-set method and passed list inclusions were employed to prune parts of the reachability tree (for detailed explanations see [5]). The objective is to minimize the sum of the delay costs. A penalty of $p(\mathbf{x}) = 0$ is passed to the *upper layer planning algorithm* (see Section 4) if all due dates are satisfied or costs $0 < p(\mathbf{x}) < p_{max}$ which are identical to the cost of delays in the objective function on the planning layer.

6. Experimental Results

The proposed two-layer approach was tested for the case study. The decisions in periods 1 to 3 are considered as first-stage decisions, those in periods 4 to 6 as second-stage decisions. The uncertainties are represented by 4 realizations of the product demands, 2 realizations of the revenues, 2 realizations of the production costs, and 2 realizations of possible breakdowns of the reactors. The realizations (B, G) of the product demands for the end-products B and G are shown in the Table below. Overall the second-stage uncertainty is represented by 32 scenarios of equal probabilities. The upper layer hybrid algorithm was implemented in MATLAB 7.3. The MILP scenario problems were for-

Realization	1	2	3	4
Period 1	(3,2)	(3,2)	(3,2)	(3,2)
Period 2	(1,3)	(1,3)	(1,3)	(1,3)
Period 3	(4,2)	(4,2)	(4,2)	(4,2)
Period 4	(2,3)	(3,3)	(3,3)	(4,1)
Period 5	(3,2)	(2,2)	(4,3)	(2,3)
Period 6	(2,3)	(1,1)	(4,4)	(1,2)

mulated using GAMS 22.5 and solved by CPLEX 10.2. In total, the scenario sub-problems consist of 8065 equations, 1249 continuous variables, and 4352 integer variables. TAOpt is employed for the optimization of the TA models in the lower layer. The global TA model is composed out of 69 TA models. It consists of 200 locations, 203 transitions, and 70 clocks. It was observed that TAOpt finds the first feasible solution after 3,000 explored nodes in most instances. TAOpt was stopped either after 20,000 nodes are explored or after a schedule with no tardiness is found. 10 runs were performed on a dual Xeon machine with 3 GHz speed. A population size of $\mu = 10$, an offspring/parents-ratio of $\lambda/\mu = 7$, and $\kappa = \infty$ were chosen. The progress plot in Fig. 2 shows the evolution of the objective function $f(\mathbf{x})$. The different lines represent the 10%-, 50%-, and 90%-quantiles. The proposed algorithm generates feasible solutions after 6 minutes. The quality of the objective is im-

proved considerably in short time, thereafter convergence slows down. In the best run, the algorithm finds the optimum solution (determined by complete enumeration of the first-stage decisions in 36 hours) in 48 minutes which has an objective function value of -21.9775. The algorithm needs about 3 sec for the evaluation of one individual without scheduling. The computation of a schedule takes up to 3.5 sec.

Fig. 2: Evolution of the two-layer algorithm

7. Conclusions

This work presents a successful application of a stage decomposition based hybrid algorithm for production planning and of the timed automata based approach for schedule generation to solve the coupled problem of production planning under uncertainty and scheduling. Ongoing work concerns the improvement of the interplay between planning and scheduling and the application to larger examples.

References

1. J. R. Birge, F. Louveaux: *Introduction to Stochastic Programming,* Springer, 1997.
2. S.L. Janak, C.A. Floudas, J. Kallrath and N. Vormbrock (2006). Production scheduling of a large-scale industrial batch plant. I. Short-term and medium-term scheduling. *Ind. Eng. Chem. Res.* 45, 8234-8252.
3. D. Wu and M. Ierapetritou (2007). Hierarchical approach for production planning and scheduling under unvertainty. *Chem. Eng. Proc.* 46, 1129-1140.
4. Y. Abdeddaiem, E. Asarin, O. Maler (2006). Scheduling with timed automata. *Theoretical Computer Science* 354, 272-300.
5. S. Panek, S. Engell, S. Subbiah, O. Stursberg (2008). Scheduling of multi-product batch plants based upon timed automata models. *Comp. Chem. Engg.* 32, 275-291.
6. N. Bauer, S. Kowalewski, G. Sand, T. Löhl (2000). A case study: Multi product batch plant for the demonstration of scheduling and control problems. *Proc*. ADPM 2000, Shaker.
7. J. Till, G. Sand, M. Urselmann, S. Engell (2007). A Hybrid Evolutionary Algorithm for Solving Two-stage Stochastic Integer Programs in Chemical Batch Scheduling. *Comp Chem Engg.* 31 , 630-647.
8. S. Subbiah,T. Tometzki, S. Engell (2008). Batch scheduling with intermediate due dates using timed automata models. *Proc*. ESCAPE 18, 151-156.

On Line Monitoring and Diagnosis of the Operation of a Hybrid CSTR by Using PCA Models

Luis G. Bergh and David Gómez

Chemical Engineering Department, Santa Maria University, Valparaiso, Chile

Abstract

Testing strategies for process monitoring, diagnosis and control is expensive, and usually requires either complex pilot plant facilities or to deal with the hard constraints posed by experimentation in real plants. The experience developing hybrid systems, combining pilot plants and models is presented. A model is used to simulate the reaction, given measured and virtual variables. Feed temperature and flow rate and reactor temperature are measured on line, while feed component concentrations are given as a stochastic disturbance. The temperature in the reactor can be controlled by regulating the cold water flow rate in the coil.

The potential of developing this class of hybrid systems is illustrated by the use of principal component analysis (PCA) methodology to build a model from experimental data. Hotelling T^2 and Q residual tests were running on line every time a steady state is reached. Both tests are used to identify the main causes of two kinds of problems: when the process target (product concentration) is not met and when a measuring device failed. Examples of these applications are discussed.

Keywords: *reactor, control, modelling, monitoring, diagnosis*

1. Introduction

In the last decade a number of techniques have been proposed to improve the benefits of chemical process operations. Some of them are: process modelling, data reconciliation, soft sensors and pattern recognition, process monitoring, fault detection and isolation, control loop monitoring, control algorithms and supervisory control. What is common to most of these areas is that the theoretical advantages of novel methods have to be confronted and tested in real plants. Simulation of the application of novel methods by using a plant model has some advantages. Once a model of a plant is available, the new methodology can be tested under a number of different conditions, at low cost. However, some disadvantages of this approach are how the model used represents the real nature of the process, the error in measurements and the disturbances coming into the plant. Most of the time, the simulations are conducted under some idealties, neglecting those aspects, and thus increasing the uncertainty and reducing the validity of results. Comparisons of different methods made on such bases may become useless, and cannot replace experimentation in the real world.

On the other hand, experimentation carry on real plants poses its own difficulties. For example, the scale factor forces that changes on some variables must be inside a narrow band, to avoid risky operating conditions or high loses in products. Some disturbances cannot always be managed, interrupting and degrading the experiments, leading to

confusing results. In summary, it is more difficult and costly to conduct a designed experience, which collected data can be properly analysed to make right decisions.

Pilot plants can reproduce most of the advantages of experimenting in a real plant, but at a small scale that relaxes some of those constraints, at the same time that provides the chance to a wide experimentation. However, expensive chemicals, large storing facilities, waste treatment units and expensive on stream analysers remain as drawbacks.

2. A different approach

Some processes can be analyzed separating the phenomena in different levels. A first level is the process hydrodynamics, heat and mass balances. A second level is the chemical change by reaction. Sometimes changes in solute concentration may also change transport properties, and the artificial decoupling will not be effective. When process hydrodynamics are not significant influenced by changes in solute concentration, then the experimentation approach on pilot plants can be simplified in two senses: Experimentally each fluid may be substituted by low cost water in a liquid phase reaction, and the solute concentration will not actually change, but an estimation can be obtained from detailed models relating measured operating conditions, such as flow rates, temperatures, pressures, levels, and concentrations.

Therefore, if a model is available, and the pilot plant is operated by using these low cost fluids, a hybrid system is developed. The real plant behaviour is simplified but the main hydrodynamics characteristics are still well represented. The variables representing the target of the process are predicted by using the model, under a wide operating region. Now, distributed control of local objectives can be administrated by supervisory control strategies based on estimation of the crucial variables. Also process monitoring, diagnosis and fault detection, isolation and remediation studies can be developed under low cost and reasonable approximation to the behaviour of a real process.

Models of different kinds can be built and solved on-line (Bergh, 2007). Here, a steady state model is used to generate the data to build a PCA model, and then this model is used to test on-line the operation of the reactor.

3. Pilot CSTR unit

A P&ID of the pilot SCTR is shown in Figure 1. Both, feed and cooling water are measured and controlled by variable speed peristaltic pumps. The electric power supplied is under control, and it is used to simulate the heat of reaction. Inlet and outlet temperatures are measured. Virtual measurement of reactant A is generated in the computer, while the outlet concentration is calculated from the model by using the measured operating variables.

A first order kinetics exothermic and irreversible reaction is considered. The input variables measured are: feed flow rate (F) and temperature (Ti), cooling water flow rate (Fw) and temperature (Twi). The output variables measured are reactor temperature (T), cooling water temperature (Two) and ambient temperature (Ta). Given an input concentration (Cao) the output concentration (Ca) is calculated from:
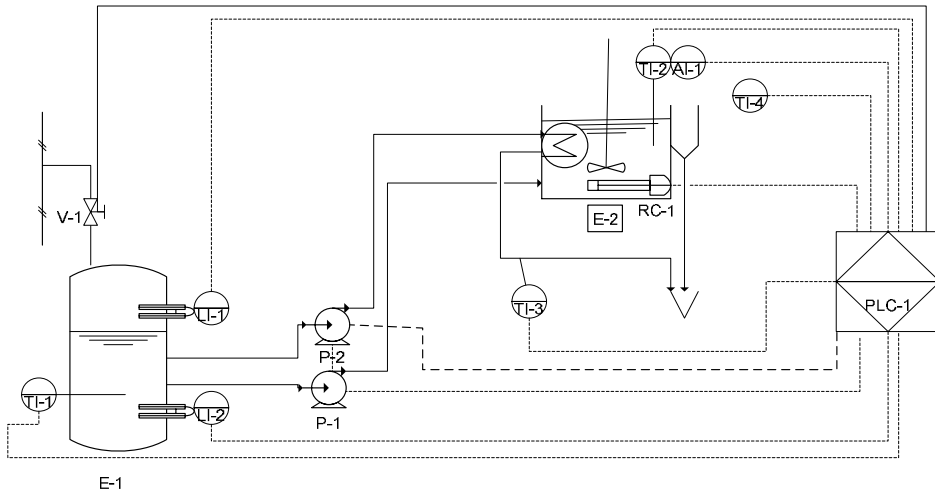
Figure 1. P&ID of pilot CSTR.

$$Ca = \frac{Cao}{(1+k\tau)} \tag{1}$$

with $k = k_0 e^{-\frac{E}{RT}}$ (2)

and $\tau = \frac{V}{F}$ (3)

The heat of reaction is calculated from: $Q = Ca \cdot k(-\Delta H)_r \cdot V$ (4)

The reactor temperature is measured at each sampling time (every 10 s), and k, Ca and Q are recalculated by using the model. The new value of heat of reaction is implemented by the electrical heater RC-1. Convergence is obtained after three or four iterations (less than one minute) and then the reactor reaches a steady state, where all material and heat balances are met. Similar iteration will occur if any of the input variable changes. This hybrid system will provide complete operating data that will be strictly valid only at steady state. In the transient period between steady states the estimation of reactor concentration will not be equal to the real one, but close enough to show smoothly the tendency. More details can be found in Gomez (2008).

4. PCA models

The key feature of PCA method is their ability to mathematically project high dimensional process and quality data into smaller dimensional, summary data sets via the development of linear models. The practical value of PCA modeling method is that this technique allows for the systematic examination and interpretation of the model outputs. Examination of the model outputs can provide insight into the operation of an industrial process during monitoring and quality assurance activities. With PCA, the systematic interpretation of dominant patterns in the data and the isolation of the most important contributors to these patterns are possible. This allows the classification of

data relationships according to normal and abnormal operation. Some of these numerous advantages of PCA method has over traditional monitoring and prediction technologies are: provision for data dimension reduction and robustness to highly correlated, noise and missing data (Kourti and MacGregor, 1995).

The concept of a latent variable model is that the true dimension of a process is not defined by the number of measured variables, but by the underlying phenomena that drive the process. The latent variables themselves are modeled as mathematical combinations of the measured variables and describe directions of variation in the original data. A latent variable model can contain many fewer dimensions than the original data, it can provide a useful simplification of large data sets, and it can allow better interpretation of the measured data during analysis (MacGregor *et al*., 2007).

5. PCA model building

Nine variables were considered, seven were measured on-line, the initial concentration was given as a date and the output concentration was calculated using heat and mass balances and kinetic model. The data were collected for different feed flow rates and feed concentration. An example of a set of data collected every 5 seconds is shown in Figure 2. When all the variables have reached a steady state, this point is a candidate to form the data matrix, after filtering. A normal operation is defined when a conversion over 95 % is reached in the reactor. More than 200 steady states were collected and filtered. This data was processed using PLS_Toolbox from Eigenvalue Research.

A PCA model was built. A model with 4 latent variables was found to explain at least 95 % of the variance in the centered and scaled pretreated data. The T^2 Hotelling, at 95 % confidence interval (using the F distribution), was 10.99 and the residuals Q limit (following Jackson, 2003), at 95 % confidence interval, was 0.29. These statistics will be used to test when the reactor operation reached an abnormal situation, and which variables are contributing the most. Two kinds of problems were studied. First, when the conversion in under 95% and when a gross error in the measurement of one operating variable occurred.

When the reactor operated with a feed flow rate of 13 mL/s, a cooling water flow rate of 20 mL/s, and a initial concentration of 1.0 mol/L, the temperature was 32.5 °C and the heat of reation was 1420 W. The conversion was 93.7% and the Q test was satisfied while the T2 Hotelling test was failed. The contribution of each variable to the failure is shown in Figure 3. The output concentration was low mainly due to a combination of high feed flow rate and high cooling water flow rate (low k and τ).

If the feed flow rate measurement is deviated by a fractional error, then the T^2 test is satisfied and the Q test is failed for errors less than -0.2 and great than 0.4, as can be observed from Figure 4. Similarly, if the input concentration measurement is deviated by a fractional error, then the T^2 test is satisfied and the Q test is failed for errors greater than 0.2.
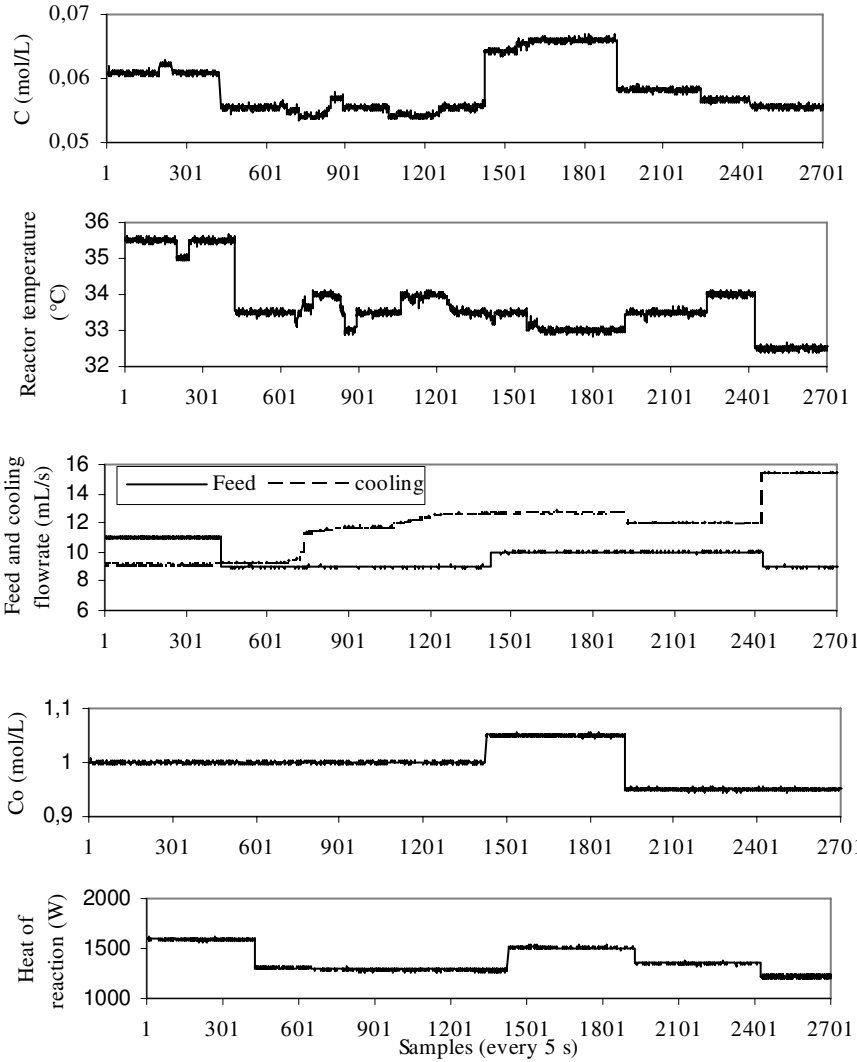
Figure 2. Operating data collected from pilot CSTR.
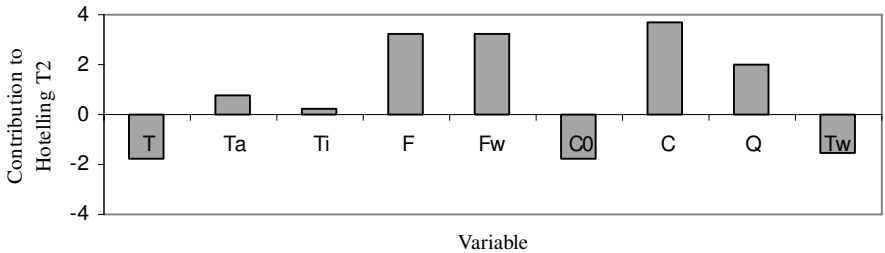
Figure 3. Contribution of variables to Hotelling's T^2.
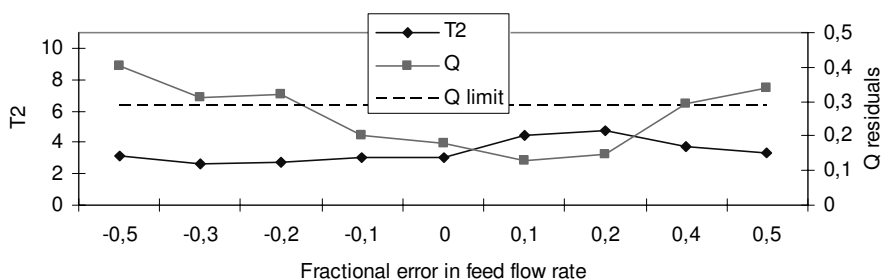
Figure 4. Sensitivity to detect fractional error in feed flow rate.

5. Conclusions

A heated and cooled tank has been operated combined with a phenomenological model to predict on-line the output concentrations of a CSTR reactor. This configuration permited to operate the reactor in a wide range, at low cost, and to test novel methodologies to detect, isolate and remediate operating and measurement failures.

The application of multivariate statistical methods, and particularly PCA, is a powerful tool to build linear models containing the essencial of the process phenomena with the minimum number of latent variables. The application of PCA models to monitoring a pilot CSTR has been demonstrated. These PCA models can be effectively used as part of a supervisory control strategy, specially when control decisions are infrequently made.

Acknowledgement

References

Bergh L.G. (2007). Combining on-line process measurements and models to empirically test strategies for process monitoring, diagnosis and control, Proceedings 12th IFAC Symposium MMM 07, Quebec City, August 21-23, 193-198.

Gómez D. (2008). Monitoreo en linea de reactor agitado calefaccionado usando métodos de proyección, Tesis magister Ing. Quím., Universidad Santa María, Valparaíso, Chile.

Jackson J.E. (2003). A user´s guide to principal components, John Wiley, New Jersey.

Kourti T. and J.F. MacGregor, (1995). Process Analysis Monitoring and Diagnosis using Multivariate Projection Methods – A Tutorial, Chemometrics and Intelligent Laboratory Systems, 28, 3-21.

MacGregor J.F., T. Kourti, J. Liu, J. Bradley, K. Dunn, H. Yu (2007). Multivariate methods for the analisys of data.bases. process monitoring, and control in the material processing industries, Proceedings 12th IFAC Symposium MMM 07, Quebec City, August 21-23, 193-198.

Using Wavelet Texture Analysis in Image-Based Classification and Statistical Process Control of Paper Surface Quality

Marco S. Reis,[a] Armin Bauer [b]

[a]*CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790, Coimbra, Portugal.*
[b]*Voith Paper Automation GmbH & Co KG, Linzer Strasse 55, 3100, St, Pölten, Austria*

Abstract

Paper formation (the level of homogeneity in the distribution of fibers on the surface of paper) is a key quality parameter for paper products, being currently monitored off-line, at low sampling rates relatively to the high production speeds achieved with modern paper machines. In this paper, we address the problem of conducting such monitoring activity on-line, in situ, using wavelet texture analysis (WTA) on raw images acquired with a especially design sensor. Our analysis shows that either a reduced set of features derived from WTA (suggested by an ANOVA analysis), or a low dimensional sub-space (with two dimensions, obtained using PCA or PLS-DA), enable an adequate separation of the several paper formation quality grades, meaning that we can indeed follow, on-line, the quality status of paper formation. Furthermore, we will also show how statistical process control (SPC) can be properly conducted using WTA features, in a simple and robust way.

Keywords: Multivariate Image Analysis, Wavelet Texture Analysis, Paper Formation, Multivariate Statistical Process Control, Principal Components Analysis.

1. Introduction

Paper formation (the level of homogeneity in the distribution of fibres on the surface of paper) is a key quality parameter for paper products, as it impacts not only the aesthetic evaluation and visual perception of quality by the end users, but also affects other relevant properties such as the average strength of paper (a sheet with poor formation is weaker than a comparable sheet with a better formation), the quality of printing operations (which is inferior for papers with poor formation) and, furthermore, leads to process problems, such as, for instance, the increase of coating consumption, as paper formation gets worse. For these reasons, paper formation constitutes a matter of great concern to paper producers, being therefore routinely measured in the quality control laboratories in modern paper production facilities, either through visual inspection and comparison with a series of standards representing different quality levels of formation or, more frequently, with resource to instrumentation that scan the paper sheet, measuring the variation of light transmission through it. This monitoring activity occurs only a few times per day, according to the routine testing plans established in each paper mill. For instance, a possible routine may consist on taken a sample for each paper reel produced, which is then sent to the control laboratory, where, after some time, the measurements are performed and results introduced in the product quality information system. This whole process introduces a very significant delay in the supervision and

control tasks, given the high production rates achieved by modern paper machines, where paper is being currently produced at linear speeds that can be over 100Km/h. Therefore, we can easily conclude that such off-line process monitoring has strong limitations, being rather ineffective in promptly detecting problems in paper formation and leading to significant losses in quality when process upsets do arise. In this context, we address in this paper the development of technology for performing paper formation monitoring, on-line and in situ, using a state-of-the-arte texture analysis methodology: Wavelet Texture Analysis (WTA). WTA enables the extraction of informative image features regarding phenomena contained in the textures (Bharati *et al.*, 2004), which are computed from the 2D wavelet transform coefficients of raw images. Wavelet analysis has been already applied to the analysis of paper formation, but only the wavelet details coefficients for the second decomposition level were used (Bouydain *et al.*, 1999), something that, according to our results, may seriously hinder the discrimination ability that can be achieved among different level of quality in paper formation.

This article is organized as follows. In the next section we briefly describe the technology underlying the image acquisition system and address the main methodological steps of our analysis approach. Then, we present some results obtained, regarding both the discrimination ability that can be achieved using WTA features, and address how can they be used for conducting statistical process control (SPC) in a proper way. Finally, we briefly conclude with a summary of the main results achieved.

2. Materials and Methods

2.1. Image acquisition sensor

The image acquisition sensor is a critical element for the on-line monitoring and analysis of paper formation. Its function is to capture good quality images in the harsh environmental conditions prevailing in the forming section of the paper machine, which is essentially steamy and wet, with drops erratically flying everywhere around the camera. The technological solution developed for the formation sensor, consists of a digital camera within a housing that is able to rotate at high speed, in order to prevent dirt accumulation on the housing surface. Such design protects the camera from the environmental conditions, ensuring that the sensor will function properly with little maintenance. The camera used was a Jai A10 CL, with a Navitar DO-2595 lens. The strobe light consists of a led array emitting red light (CCS LDL-TP).

2.2. Methods
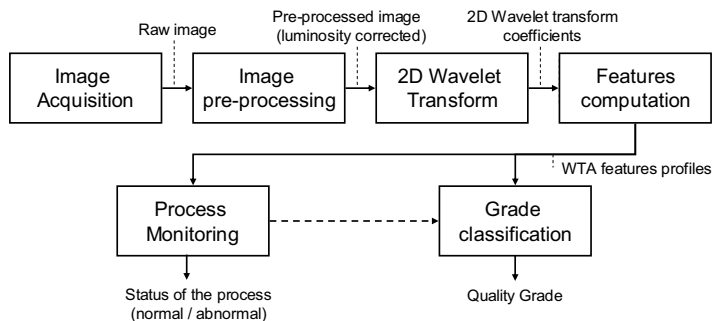
Figure 1. Scheme of the WTA methodology for monitoring and classification of paper formation.

The basic methodological steps considered in our analysis are as follows (Figure 1). Images are first acquired (i) and corrected for non-uniform illumination effects (ii), even

though this is not a critical step in the analysis, as our procedure is rather robust to this kind of problems (i.e., from our experience, final results do not change much upon removing such pre-processing task from the analysis). Then, we (iii) computed the 2D wavelet transform for each raw image (Mallat, 1999), which essentially consists of transforming a matrix of numbers (pixel intensities, as we are analyzing single-channel or grey-level images) into another, with the same size (same overall number of wavelet coefficients), containing blocks of coefficients regarding details for different scales (from the finest to the coarsest scale, which is known as the decomposition depth, J_{dec}) and along three different directions (horizontal, vertical and diagonal). In our case, the decomposition depth considered was $J_{dec} = 5$, and therefore each image will give rise to 16 blocks of data or sub-matrices ($=J_{dec}$x3+1; the "1" stands for the block of final approximation coefficients, containing the lowest resolution version of the image, after removing all the details corresponding to the finer scales). The next stage in WTA is (iv) to summarize the information contained in the wavelet coefficients in each of these blocks into a single number, from which a vector of WTA features will result. From the several possibilities available (energy, entropy, averaged l_1-norm, standard deviation) we used the standard deviation (the wavelet transform of uncorrelated homogeneous white noise is also white noise with the same standard deviation; therefore by using the standard deviation to summarize the information in each block, we are tactilely using a stochastic uncorrelated frame of reference against which we compare the texture in the images). As the end of this fourth stage, we have compressed the original analysis space composed by 512x512 pixels or intensity numbers, to just 16 numbers that are expected to contain the essential information necessary to discriminate different formation patterns. Finally, the extracted WTA features can be used for (v) classifying the acquired image by proposing a quality grade and/or for conducting statistical process control of this property. The first activity (classification) requires the development of a suitable classifier, which usually involves a feature selection/compression stage and a subsequent classification stage (Pal and Mitra, 2004). For the first task we have considered here variable selection (using ANOVA) as a feature selection methodology and partial least squares for discriminant analysis (PLS-DA), as well principal components analysis (PCA), for features compression. As to the second task (process monitoring), we applied multivariate statistical process control (MSPC) based on PCA, in order to deal with the high collinearity present in the WTA features (Jackson and Mudholkar, 1979; Kresta *et al.*, 1991).

3. Results

A set of 24 images, representing different paper formation patterns, was collected for analysis. These raw images were classified, by visual inspection, into three quality levels regarding paper formation: 1-"good", 2-"poor", 3-"bad" (Figure 2). For two images were given special annotations: a dubious sample was assigned the class label "1,5", and an outlier, possibly resulting from an image acquisition problem, was labeled with score "9". Following the sequence of steps represented in Figure 1, these images were pre-processed for non-uniform illumination correction and transformed to 2D wavelet coefficients. Finally, the wavelet feature profiles were computed, containing 16 features for each image, based on which the subsequent analysis was performed. This analysis had two main components: assessing the discrimination ability of wavelet features to discern among different formation quality grades (classification) and to study how statistical process control could be conducted based on such features (monitoring).
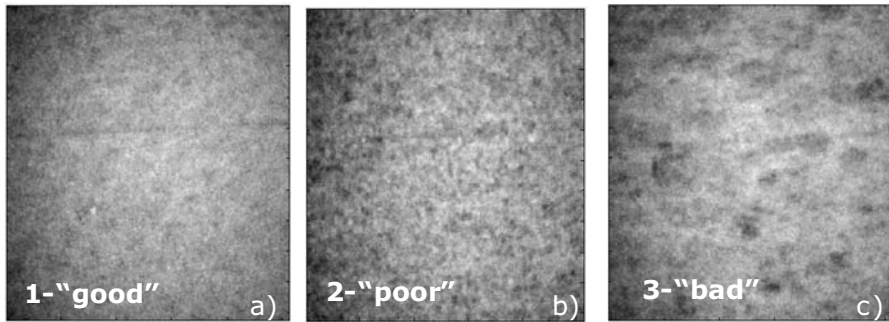
Figure 2. Images representative of the three main quality grades of paper formation.

3.1. Classification of paper formation

Even though the compression magnitude of the entire image information to 16 features is indeed quite large, such features are still highly redundant. For instance, the first principal component (PC) in a PCA analysis over these 16 features, represents 97.96% of the overall data variability. However, in order to achieve a proper discrimination, a minimum of two PC must be retained (we get to this number by computing the error rate estimated by "leave-one-out" cross-validation using an increasing number of retained PC, and applying the simple linear classifier to perform the classification task, from which we verified that, after retaining the second PC, the error rate dropped from ~ 23% to 0%). Figure 3 illustrates the separation ability obtained from different compression strategies (to reduce the number of features from 16 to 2), followed by the estimation of a linear classifier in the resulting reduced map (as the separation is so good, there is no need to use more complex classifiers): PCA, PLS-DA (Barker and Rayens, 2003) and variable selection (using ANOVA to select those features that most descriminate the various formation quality levels), followed by Fisher Discriminant Analysis (FDA), as this method alone is not appropriate to handle situations were the classifying features are heavily correlated, due to matrix inversion problems, as happens in this case. As we can verify, all methods successfully discriminate among the three levels of quality present in the sample images, meaning that WTA does provide an adequate basis for analyzing paper formation.
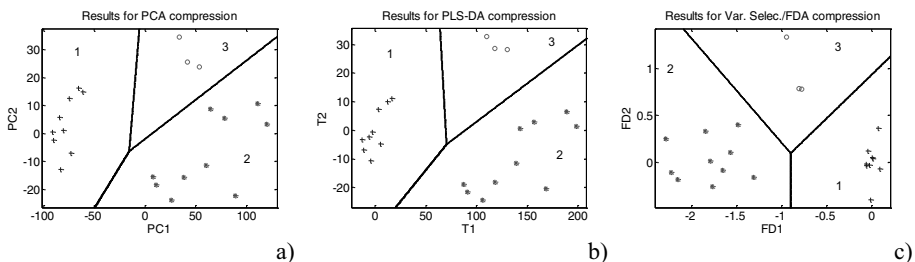
Figure 3. Results for the classification of samples after compressing the number of WTA features from 16 to 2: a) using PCA; b) using PLS-DA; c) using variable selection (based on ANOVA), where 3 variables were selected, and FDA.

3.2. Monitoring paper formation

In order to use the WTA feature profiles for monitoring purposes, a suitable statistical process control scheme was adopted, with well established results and adequate for

addressing the highly correlated nature of the data matrix of wavelet features: multivariate statistical process monitoring (MSPC) based on PCA (PCA-MSPC) (Jackson, 1959; Kresta *et al.*, 1991). This methodology consists of modelling systems operation under normal conditions using a PCA model, in order to capture the normal correlation structure and variability relative to the so called "common" causes. For each (new) multivariate observation (x_i), two statistics are computed, one relative to the (Mahalanobis) distance from its projection onto the PCA subspace to the centroid of the normal operation conditions (NOC) cluster, also in the PCA subspace, usually referred as T_i^2, and another relative to the (squared) Euclidean distance to such a subspace, Q_i.

In order to assess the application of this approach to the on-line monitoring of paper formation, a sub-set of four images from class 1-"good" were selected, representing normal operation conditions (NOC; images 1, 2, 7 and 13), from which the method parameters, including the statistical control limits, were estimated. All the remaining images, with representatives from all grades of quality, were used for testing the monitoring procedure. Figure 4 depicts the projection of all samples into the PCA subspace estimated from the wavelet features for the four reference images, where it is clear that all reference and testing samples from normal operation conditions fall inside the NOC region (95% control ellipse, also computed only using the reference set), whereas abnormal samples are also correctly signalled as such, falling outside the NOC ellipse.
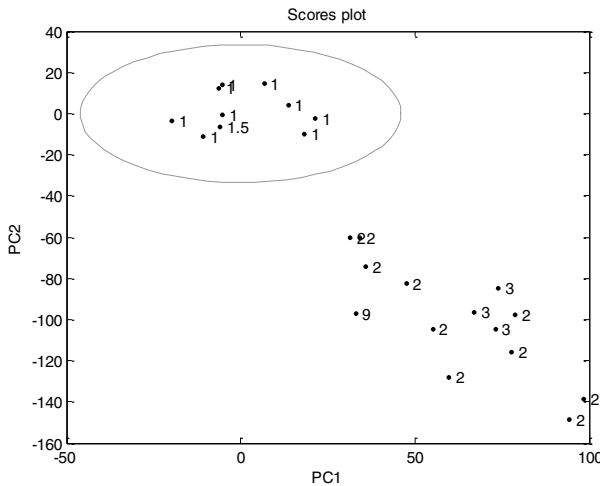
Figure 4. Projection of all samples (train and test) on the PCA subspace estimated from the four training images of quality grade 1-"good", taken as representative of NOC, with the associated 95% NOC control ellipse. All abnormal samples in the test set (class 2 or 3) fall outside the control ellipse.

From Figure 4, it is possible to see that the dubious sample labelled "1.5" also falls inside the NOC region, which is not unexpected given its similarity to other images from this class. However, an analysis of the method statistics (T^2 and Q) clarify its status and rather abnormal texture (see plot for Q), indicating that it is indeed an abnormal image (Figure 5; note the logarithmic scale used in the YY' axis, given the large magnitude difference between the statistics values for the normal and abnormal situations).
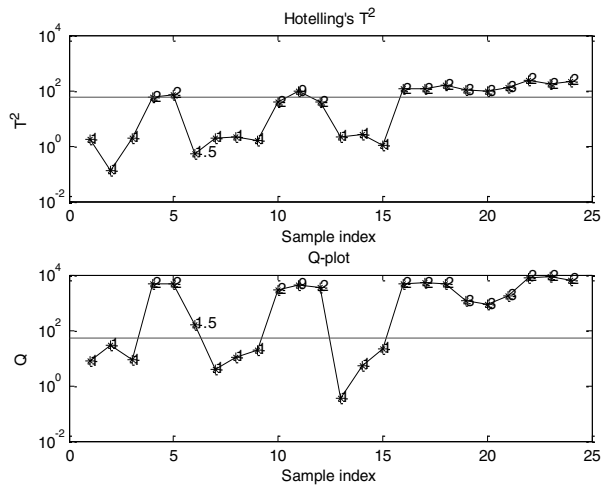
Figure 5. Control charts for the T^2 and Q statistics, regarding the implementation of PCA-MSPC over all images (train and test). Using the combination of the two statistics, all samples are correctly classified as normal or abnormal. (Note the logarithmic scale used in the plots).

4. Conclusions

In this article, we have presented an approach for classifying the quality grade of paper formation based on the WTA features profiles computed from images taken on-line from a pilot scale paper machine. Furthermore, their application to process monitoring was also addressed. Our analysis showed that the classification task is successfully conducted in a reduced dimensional space (2 dimensions), using either a reduced set of features (suggested by an ANOVA analysis), or a low dimensional sub-space (obtained through PCA or PLS-DA). Regarding the monitoring task, it was also possible to verify that statistical process control (SPC) can be properly conducted using WTA, in a simple and robust way. Therefore, we can conclude from the results presented, that it is possible to simultaneous monitor and assess formation quality, on-line and in-situ, using WTA-analysis of images acquired with the special designed sensor developed.

References

Barker, M. & Rayens, W., 2003, Partial Least Squares for Descrimination, *Journal of Chemometrics, 17*, 166-173.

Bharati, M. H., Liu, J. J. & MacGregor, J. F., 2004, Image Texture Analysis: Methods and Comparisons, *Chemometrics and Intelligent Laboratory Systems, 72*, 57-71.

Bouydain, M., Colom, J. F. & Pladellorens, J., 1999, Using Wavelets to Determine Paper Formation by Light Transmission Image Analysis, *Tappi Journal, 82*, 7, 153-158.

Jackson, J. E., 1959, Quality Control Methods for Several Related Variables, *Technometrics, 1*, 4, 359-377.

Jackson, J. E. & Mudholkar, G. S., 1979, Control Procedures for Residuals Associated with Principal Component Analysis, *Technometrics, 21*, 3, 341-349.

Kresta, J. V., MacGregor, J. F. & Marlin, T. E., 1991, Multivariate Statistical Monitoring of Process Operating Performance, *The Canadian Journal of Chemical Engineering, 69*, 35-47.

Mallat, S., 1999, *A Wavelet Tour of Signal Processing* (2nd ed.). San Diego: Academic Press.

Pal, S. K. & Mitra, M., 2004, *Pattern Recognition Algorithms for Data Mining*. Boca Raton: Chapman & Hall/CRC.

Numerical Pitfalls by State Covariance Computation

Nina P. G. Salau,[a] Jorge O. Trierweiler,[a] Argimiro R. Secchi,[b]

[a]*GIMSCOP/DEQUI/UFRGS, Campus Central, Porto Alegre/RS, CEP 90040-040, Brazil*
[b]*PEQ/COPPE/UFRJ, CP 68502, Rio de Janeiro/RJ, 21945-970, Brazil*

Abstract

With the error past information, conveyed by state covariance matrix, a state estimator can predict the state vector one step ahead, thus providing system operators with more information to make control decisions such as security assessment and other related functions. A well-computed state covariance matrix avoids error propagation due to numerical pitfalls and, thereby, is crucial for a successful state estimator design. In this paper we investigate the numerical robustness of four EKF formulations and derive some results giving insights into their numerical performance. As benchmark case study, we have chosen a batch chemical reactor with reversible reactions whose system model and measurement are such that multiple states satisfy the equilibrium condition. Finally, we have shown that CEKF is the best alternative to EKF for such case study.

Keywords: nonlinear state estimation, state covariance, multiple solutions.

1. Introduction

It is well known that a suitable design of state estimators requires a representative model for capturing the plant behavior and knowledge about the noise statistics, which are generally not known in practical applications (Vallapil & Georgakis, 2000). In practical applications, however, some pitfalls such as numerical round-off, plant-model mismatch and state unobservability deserve also special attention because they can lead to divergence problems (Brown & Hwang, 1996). Any state covariance matrix equation is composed by states and measurements linear models and noise covariance statistics and, hence, all of the mentioned pitfalls may increase the error propagation conveyed by this matrix. In the literature, several modified implementations of the EKF are presented in an effort to avoid numerical pitfalls (Simon, 2006). The basic difference between these formulations is concerned with the state covariance matrix computation. In this work we investigate the numerical robustness of four EKF formulations applied to a system tending to an equilibrium condition. Firstly, we outline a condition that leads to a classical EKF formulation to converge to physically unrealizable state and demonstrate the potentiality of two other EKF formulations to handle with this numerical pitfall. Next, we demonstrate that the constrained extended Kalman filter (CEKF) is the best alternative to EKF for such system due to the possibility of incorporating constraints into an optimization problem preventing, hence, the estimator from converging to physically unrealizable states.

3. Formulation and Solution of the Estimation Problem

Consider the following nonlinear dynamic system to be used in the state estimators of this work:

$$\dot{x} = f\left(x, u, t\right) + \omega\left(t\right) \tag{1}$$

$$x_k = f_{k-1}\left(x_{k-1}, u_{k-1}, t_k\right) + \omega_{k-1} \tag{2}$$

$$y_k = h_k\left(x_k, t_k\right) + v_k \tag{3}$$

where u denotes the deterministic inputs, x denotes the states, and y denotes the measurements. The process-noise vector ω (ω_k) and the measurement-noise vector v_k are assumed to be white Gaussian random processes with zero mean and covariance Q (Q_k) and R_k, respectively.

1.1. Unconstrained EKF formulations

We have selected three different unconstrained EKF formulations to be compared in our work: a classical EKF formulation known as discrete EKF (DEKF), the EKF with continuous Riccati equation (EKF-CRE) and the reduced-rank extended Kalman filter (RREKF). The basic difference of these formulations is related to state covariance matrix computation, as given in Table 1.

Table 1. Three unconstrained formulations of EKF

Filter	State Covariance Matrix		
DEKF	Transition	$P_{k\|k-1} = F_{k-1}P_{k-1\|k-1}F_{k-1}^T + Q_{k-1}$	(4)
	Update	$P_{k\|k} = \left[I_n - K_k H_k\right]P_{k\|k-1}$	(5)
EKF-CRE	Transition & Update	$P_{k\|k} = P_{k-1\|k-1} + \int_{k-1}^{k} \begin{bmatrix} F(\tau)P(\tau) + P(\tau)F(\tau)^T + Q(\tau) \\ -P(\tau)H(\tau)^T R(\tau)^{-1} H(\tau)P(\tau) \end{bmatrix} d\tau$	(6)
RREKF	Transition	$P_{k\|k-1} = F_{k-1}P_{k-1\|k-1}F_{k-1}^T + Q_{k-1}$	(7)
		$P_{q,k\|k-1}^{rr} = \sum_{i=1}^{q} \lambda_{P,i}\upsilon_{P,i}\upsilon_{P,i}^T$	(8)
	Update	$P_{k\|k} = \left[I_n - K_k H_k\right]P_{k\|k-1}$	(9)

1.2. Moving Horizon Estimator

MHE is described by the following optimization problem (Muske & Rawlings, 1994; Rao & Rawlings, 2003).

$$\min_{\substack{\omega_{k-N-1\|k},\cdots,\omega_{k-1\|k} \\ v_{k-N\|k},\cdots,v_{k\|k}}} \Psi_k^N = \begin{bmatrix} \hat{\omega}_{k-N-1\|k}^T \left(P_{k-N-1\|k-1}\right)^{-1} \hat{\omega}_{k-N-1\|k} \\ + \sum_{j=k-N}^{k-1} \hat{\omega}_{j\|k}^T \left(Q_{k-1}\right)^{-1} \hat{\omega}_{j\|k} + \sum_{j=k-N}^{k} \hat{v}_{j\|k}^T \left(R_k\right)^{-1} \hat{v}_{j\|k} \end{bmatrix} \tag{10}$$

The state covariance matrix is computed by the Riccati discrete equation (Rao & Rawlings, 2003).

1.3. Constrained EKF

CEKF follows from MHE for a zero-length horizon ($N=0$). If the measurement equation is linear, the resulting problem is a quadratic program which can be solved with small computational effort (Gesthuisen et al., 2001). CEKF is similar to EKF because also takes into account a zero-length horizon in the updating stage.

2. Example of Numerical Pitfalls by State Covariance Computation

2.1. Case Study

As benchmark case study, we have chosen a batch reactor process, as introduced by Haseltine & Rawlings (2005). All reactions are reversible and follow the ideal gas law:

$$A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} B + C$$

$$2B \underset{k_4}{\overset{k_3}{\rightleftharpoons}} C \tag{11}$$

Multiple states satisfy the equilibrium condition for a given measurement, which in this case is the system pressure at the equilibrium, evaluated by the following equation:

$$y = p = (c_A + c_B + c_C) RT \tag{12}$$

Table 2 presents the possible theoretical solutions, without measurement or state noise, at the equilibrium pressure given by the initial state: $c_0 = \begin{bmatrix} 0.5 & 0.05 & 0 \end{bmatrix}$. Note that only the solution EQ1 has physically realizable states (nonnegative concentrations).

Table 2. Equilibrium points

Component	EQ1	EQ2	EQ3
c_A	0.0122	-0.0267	-1967.4
c_B	0.1826	-0.2372	-9.9454
c_C	0.6669	1.1257	1978.2

The state estimation parameters and the poor initial guess $\bar{x}_0$ used for this example were obtained from Haseltine & Rawlings (2005):

$$\Delta t = t_k - t_{k-1} = 0.25 \ min \tag{13}$$

$$P_0 = diag \begin{pmatrix} 0.5^2 & 0.5^2 & 0.5^2 \end{pmatrix} \tag{14}$$

$$Q = diag \begin{pmatrix} 0.001^2 & 0.001^2 & 0.001^2 \end{pmatrix} \tag{15}$$

$$R = 0.25^2 \tag{16}$$

$$\bar{x}_0 = \begin{bmatrix} 0 & 0 & 4 \end{bmatrix} \tag{17}$$

According to the authors, EKF fails in this example because the system model and measurement are such that multiple states satisfy the equilibrium condition and is given a poor initial guess of the state for the estimator. Nonetheless, we cannot assert that EKF fails because two equilibrium points (EQ1 and EQ2 in Table 2) are possible due to the poor and incoherent initial state and its respective state covariance initial guess used in this example. Further, the theory of EKF formulations assumes system observability, i.e., the state dynamic system is uniquely determinable from its inputs and outputs, given a model for the dynamic system. Since at least two states exist that realize the same equilibrium pressure, the batch reactor system is considered unobservable according to the nonlinear observabilty. Because this system satisfies the observability rank condition (linear observabilty), it is also considered locally weakly observable (Hermann and Krener, 1977). As the nonlinear observability assumption is violated, it is not surprising that the theory of EKF formulations may not work for this example. Details on observability definition can be found in Sontag (1998).

2.2. Results and Analyses

2.2.1. Measurement Noise Perturbation

Firstly, uniform and normally distributed random measurement noises were simulated. Either solution EQ1 or solution EQ2 is obtained in accordance with the set of random measurement noise employed into DEKF. Thus, we have chosen a set of random measurement noise which leads DEKF to converge to the solution EQ1 and added a noise measurement perturbation in order to lead it to converge to solution EQ2. As it

can be seen in Fig. 1, a noise measurement perturbation of 0.754 *atm* at *t* = 0.5 *min* changes the estimated states trajectory from solution EQ1 to solution EQ2.

(a)   (b)
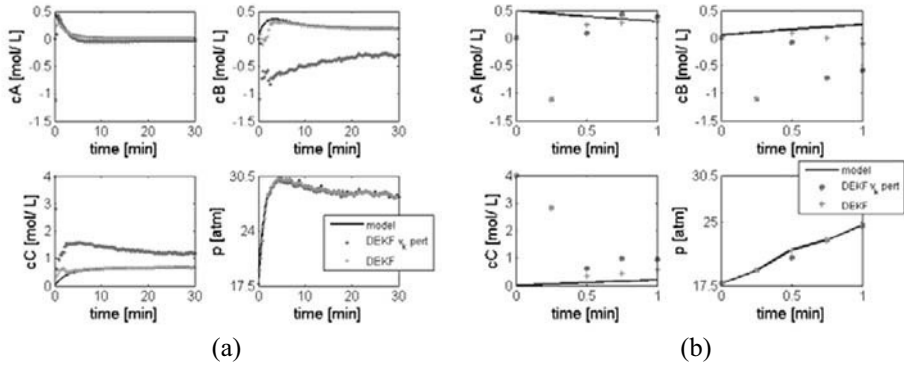
Fig. 1. Comparison between the performances of DEKF with and without a noise measurement perturbation of 0.754 *atm* at *t*=0.5 *min*: (a) until final batch time and (b) until t=1 min.

We have also experienced that other numerical pitfalls can lead DEKF to converge to solution EQ2, such as low-precision arithmetic, gradient (linear model) calculation by finite differences, poor guess of P_0 and incorrect values of tuning parameters Q and R.

2.2.2. Comparison between Unconstrained EKF formulations

To prevent the state estimator from converging to EQ2, two other unconstrained EKF formulations were applied to this example. In spite of a measurement noise perturbation, RREKF and EKF-CRE have converged to solution EQ1, as shown in Fig. 2(a).

(a)   (b)
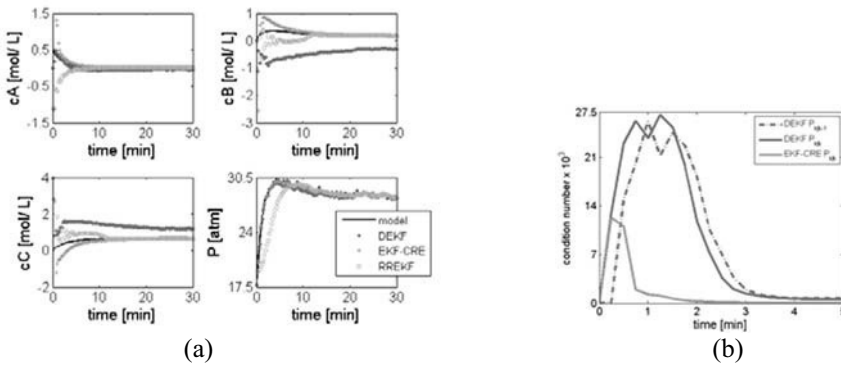
Fig. 2. Comparison between (a) DEKF, EKF-CRE and RREKF performances; (b) the condition number of DEKF and EKF-CRE state covariance matrices.

RREKF disregards non-dominant eigenvectors which implies zero variance in the respective directions and no effect of measurement updates. For this example, the non-dominant eigenvector is orthogonal to the tangent of solution EQ1 (the scalar product approaches zero). RREKF converges to EQ1 because it applies no correction in the non-dominant eigenvector direction.

EKF-CRE is not subjected to errors due to model discretization and the term $PH^T R^{-1} HP$ inserts nonlinearity into the respective state covariance matrix. Besides, the state covariance matrix computed by EKF-CRE in a single stage (Eq. 6) presents a smaller condition number, i.e., it is less sensitive to perturbations than the states

covariance matrices computed by DEKF (Eqs. 4 and 5), as shown in Fig. 2(b). The mentioned advantages of EKF-CRE over DEKF justify the convergence of this formulation to EQ1, even with a measurement noise perturbation.

Although EKF-CRE and RREKF have prevented from physically unrealizable states at the final batch time, physically unrealizable states (negative concentrations) were unavoidable during the batch. This fact is justified by the poor state and respective state covariance initial guesses.

2.2.3. Comparison between clipped DEKF and CEKF

To prevent physically unrealizable states, we have constrained DEKF with an ad hoc clipping strategy in which negative update values of the state are set to zero (i.e., if $\hat{x}_{k|k} < 0$, set $\hat{x}_{k|k} = 0$). The comparison between clipped DEKF and CEKF performances are shown in Fig. 3(a). Before eventually converging to the actual states, the pressure filtered by the clipped DEKF is quite larger than the measured one. Due to a lower initial guess of the state covariance matrix, $P_0 = diag\begin{pmatrix} 0.022^2 & 0.022^2 & 0.022^2 \end{pmatrix}$, the clipped

DEKF in our work has presented a better performance in comparison to performance of the clipped DEKF in Haseltine & Rawlings (2005). In their paper, the clipped DEKF drives the predicted pressure 3 orders of magnitude larger than the measured pressure before eventually converging to the actual states at a longer time scale (1 order larger than the converge time obtained in our work). A clipped DEKF, however, disregards the assumption that v_k is a Gaussian random noise and does not let the Kalman gain to distribute properly the measurement residual throughout the estimated state and, thereby, corrects them. On the other side, CEKF swiftly converges to the actual states and minimizes ω_k and v_k in a least square sense, incorporating constraints into an optimization problem, which prevents from bad noises distribution. Fig. 3(b) illustrates the distribution of v_k for the clipped DEKF and the CEKF.

(a)   (b)
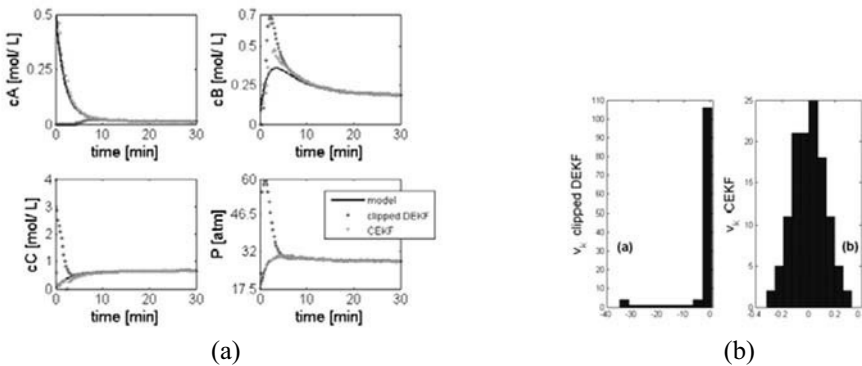
Fig. 3. Comparison between (a) clipped DEKF and CEKF performances and (b) the distribution of v_k for clipped DEKF (non-Gaussian) and CEKF (Gaussian).

Finally, we have compared the performances between CEKF and MHE (N=2). The high computational effort of MHE (2 orders larger) is not justified for this process as the improvement of estimation results is marginal, as shown in Fig. 4.

Due to the small computational effort of CEKF, possibility of incorporating constraints into an optimization problem, and the comparable estimation results to MHE, this approach seems to be the best choice for our case study.
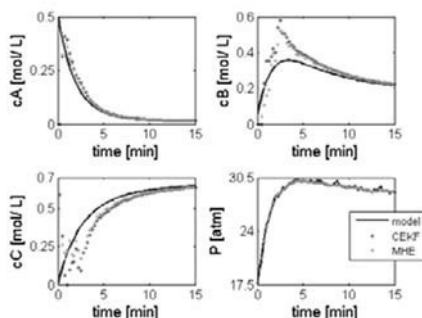
Fig.4. Comparison between CEKF and MHE (*N*=2) performances.

3. Conclusions

This paper outlines the numerical robustness of four EKF formulations for a chemical engineering example whose system model and measurement are such that multiple states satisfy the equilibrium condition. With a measurement noise perturbation, we outline a condition that has led a classical EKF formulation (DEKF) to converge to physically unrealizable state. According to our results, EKF-CRE and RREKF are more numerically robust in computing the state covariance than the DEKF. As both formulations have avoided an increase in error propagation due to a measurement noise perturbation, they were able to converge to the desired final state solution. Thus, a suited choice of the EKF formulation based on the state covariance equation is essential to prevent from physically unrealizable states. However, due to the poor initial states and its respective state covariance initial guess for the batch reactor, EKF-CRE and RREKF have not prevented from physically unrealizable states during the batch. Hence, CEKF can be seen as the best alternative to systems with such behavior due to the possibility of incorporating constraints into an optimization problem to prevent from physically unrealizable states. Besides, it is simpler, computationally less demanding than the MHE, and has comparable performance.

References

Brown, R.G.; Hwang, P.Y.C (1996). Introduction to Random Signals and Applied Kalman Filtering: With MATLAB Exercises and Solutions. John Wiley & Sons, U.S.A.

Gesthuisen, R.; Klatt, K. -U.; Engell, S. (2001). Optimization-based State Estimation: a Comparative study for the Batch Polycondensation. In Proceedings of ECC-2001.

Haseltine, E.L.; Rawlings, J.B. (2005). Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation. Ind. Eng. Chem. Res., 44, 2451-2460.

Hermann, R. and A. Krener (1977). Nonlinear controllability and observability. Automatic Control, IEEE Transactions, 22(5), 728–740.

Muske, K.R.; Rawlings, J.B. (1994). Nonlinear Moving Horizon State Estimation. NATO ASI Series, Kluwer Academic, 293, 349–365.

Rao, C.V.; Rawling, J.B.; Mayne, D.Q. (2003). Constrained State Estimation for Nonlinear Discrete-Time Systems: Stability and Moving Horizon Approximations. IEEE Trans. Autom. Control, 48(2), 246-258.

Simon, D. (2006). Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches. John Wiley & Sons, U.S.A.

Sontag, E.D. (1998). Mathematical Control Theory: Deterministic Finite Dimensional Systems. 2nd Edition, Springer, New York.

Valappil, J.; Georgakis, C. (2000). Systematic Estimation of State Noise Statistics for Extended Kalman Filters. AIChE J., 46(2), 292-398.

1221

IPL2&3 Performance Improvement Method for Process Safety Using the Event Correlation Analysis

Junya Nishiguchi, Tsutomu Takai

Yamatake Corporation, 1-12-2, Kawana, Fujisawa-shi, Kanagawa, Japan

Abstract

Alarm management efforts have recently intensified in the world, and there are many tools based on a guideline EEMUA191 to evaluate alarm system performance. However, the improvement and rationalization of alarm systems have not made satisfactory progress, because these tools usually just provide with alarm system performance metrics but not useful information to improve and rationalize alarm systems. In this paper, a novel method for performance improvement of IPL2&3 including alarm system using the event correlation analysis was proposed.

Keywords: EEMUA, IPL, alarm management, event correlation, clustering

1. Introduction

Safe operation is the top priority for process plants. As a concept of safe design to provide protection from hazardous incidents, Independent Protection Layers (IPLs) which consist of eight layers as shown in Table 1 has been extensively applied to the various plants (AIChE/CCPS 1993). The second layer (IPL2) and third layer (IPL3) are related to alarm system. The primary purpose of IPL2 is the supervision of the plant under the normal operation with Basic Process Control System. When the process variables deviate from the set points, the system activates alarms and requires the corrective actions of the operator. On the other hand, IPL3 represents the critical alarms and corresponding operator interventions. Failure in the functions of IPL2&3 is resulting to production loss. Therefore, it is important that IPL2&3 should be able to function effectively from both safety and production standpoint.

Table 1. Independent Protection Layers for Process Safety

Layers	Definitions	Functions	
IPL8	Community Emergency Response	Minimize damage from an incident	
IPL7	Plant Emergency Response		
IPL6	Physical Protection (Dikes)		
IPL5	Physical Protection (Relief Devices)		
IPL4	Automatic Action SIS or ESD		Prevent an incident from happening
IPL3	Critical Alarms, Operator Supervision, and Manual Intervention	Ensure safe and productive operations	
IPL2	Basic Controls, Process Alarms, and Operator Supervision		
IPL1	Process Design		

For rationalization of alarm system, EEMUA191 (EEMUA 1999) is widely accepted as a de facto standard guideline. In EEMUA191, alarm system is defined as *a very important way of automatically monitoring the plant condition and attracting the attention of the process plant operator to significant change that require assessment or*

action. It specifies several performance metrics that can be used to assess the alarm system performance:

- ・ *Operator questionnaires*
- ・ *Alarm usefulness surveys*
- ・ *Assessment of number of alarms in a system*
- ・ *Measurement of average alarm rate*
- ・ *Measurement of number of alarms following a major plant upset*
- ・ *Measurement of operator response time*
- ・ *Measurement of number of standing alarms*
- ・ *Analysis of the priority distribution of alarms configured and occurring*
- ・ *Correlation techniques*

Unfortunately the effective correlation techniques from the standpoint of rationalizing IPL2&3 had not been defined well. Analyzing correlation between the alarms and operator actions in a process upset is expected to estimate the propagation path, source origin, and nuisance alarms, which can be useful information to improve the alarm system and understand the IPL2&3 performance properly.

2. New Method for IPL2&3 Performance Improvement

2.1. Concept

As referred to in EEMUA191, every alarm should be defined uniquely to notify a specific upset situation, and every operator action should be defined to solve the corresponding alarm. Once a certain upset causes alarms in the plant with the poor IPL2&3, it will propagate via the process fluid and consequently leads to the alarm flood; meanwhile, alarms without the corrective actions will increase operators' workload as nuisance alarms. Therefore, in order to rationalize IPL2&3 performance, every alarm and operator action should be properly related each other.

As a conventional method for fault propagation analysis, Signed Directed Graphs (Shiozaki et al. 1989) has been actively researched. However, it is difficult for this method to put into practical use because it requires much effort to adjust process model whenever the process devices are replaced. Also, to our knowledge there has not been any systematic method for rationalizing operator actions.

The concept of our method is to provide a fast, easy, and effective way to improve IPL2&3 performance. The relationship and occurrence order between events, which contain alarm and operator action events, are extracted from event log data with Event Correlation Analysis (Nishiguchi et al. 2005). In this method, the event pairs with constant occurrence time lags are considered to be similar events, since the time lags depend on the time delay of the process dynamics and human reaction time. From the similarity and the occurrence time lag calculated by the Event Correlation Analysis, we can estimate the group with the consequential events as well as their occurrence order. Our method enables us to extract the relationships between process variables change and operators' behavior only with accumulated log data. The solutions to improvement for the IPL2&3 design and management are found rapidly and easily. The examples of extracted relationships and corresponding solutions are as follows.

I. Consequential Alarms

When the several alarms are strongly related, these alarms are likely to be consequential alarms, which can be reduced by alarm filtering techniques.

II. Complex Operator Actions

When the several operator actions are strongly related, the operator actions are likely to be complex sequential operations, which can be reduced by automating operations.

III. Redundant Alarms

When there is a high frequent alarm without the related operator actions, it may be redundant alarm, which can be reduced by changing set point or replacing into message.

IV. Causes of upset

When constant occurrence order of alarms and operator actions is found, the first occurrence event is likely to be the source of plant upset.

2.2. Event Correlation Analysis

Event Correlation Analysis quantifies the relationship and occurrence order between alarms and operator actions. Although correlation coefficient is usually used to measure the relationship between two continuous values, it is well known fact that the coefficient cannot be applied for event data (Li 1990), such as alarms and operator actions log data. In this method, similarities between all event pairs are calculated from the log data with probability distribution of correlation regarding independent event pairs (Figure 2).

The alarms and operator actions log data obtained from Distributed Control System (DCS) contains occurrence time and event kind. The log data is converted into event time series $s_i(t)$ defined as binary series for each event i, which is 1 if the event occurs within the time window Δt and 0 if does not.

$$s_i(t) = \begin{cases} 1, & \text{if some points in } (n\Delta t, (n+1)\Delta t] \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where Δt is window size, and n is time unit.

The cross correlation function (2) between event i and j indicates the occurrence counts that event j follows event i in the time range of $(m\Delta t, (m+1)\Delta t]$.

$$c_{ij}(m) = \begin{cases} \sum_{n=1}^{T/\Delta t - m} s_i(n)s_j(n+m) & m \geq 0 \\ c_{ji}(-m) & m < 0 \end{cases}, \quad -K \leq m \leq K \tag{2}$$

where T is observation time, and K is maximum lags of time unit.

From (2) maximum correlation value c^*_{ij} and time lag at maximum correlation m^*_{ij} are defined as follows.

$$c^*_{ij} = \max c_{ij}(m), \quad m^*_{ij} = \arg\max_m c_{ij}(m) \tag{3}$$

As a result, the similarity between two events R_{ij} is defined with the probability that the correlation between two independent events is lower than c^*_{ij} within time lag K.

$$R_{ij} = P\left(c_{ij}(m) < c^*_{ij} \mid -K \leq m \leq K\right) \cong \left\{ \sum_{l=0}^{c^*_{ij}-1} \frac{v^l e^{-v}}{l!} \right\}^{2K+1} \tag{4}$$

where v is the expected value of Poisson distribution regarding the occurrence of independent events, which is approximated by the average co-occurrence number between independent two events as shown in equation (5).

$$v = \frac{T}{\Delta t} \cdot p_i \cdot p_j \cong \frac{\Delta t}{T} \sum_{n=0}^{T/\Delta t} s_i(n) \cdot \sum_{n=0}^{T/\Delta t} s_j(n) \tag{5}$$

In other word, this method conducts the statistical test of hypothesis that two events i and j are generated independently. The actual maximum correlation c^*_{ij} is compared

with the distribution of correlation value between independent two events. The similarity R_{ij} is defined with the value, which is subtracted the reject rate from one.

By applying some clustering algorithm (e.g. hierarchical clustering) with the similarities of every event pair, the event groups with similar events will be extracted automatically. In addition, the occurrence order of each event pair is estimated with the time lag m^*_{ij}, which represents the most probable time delay.
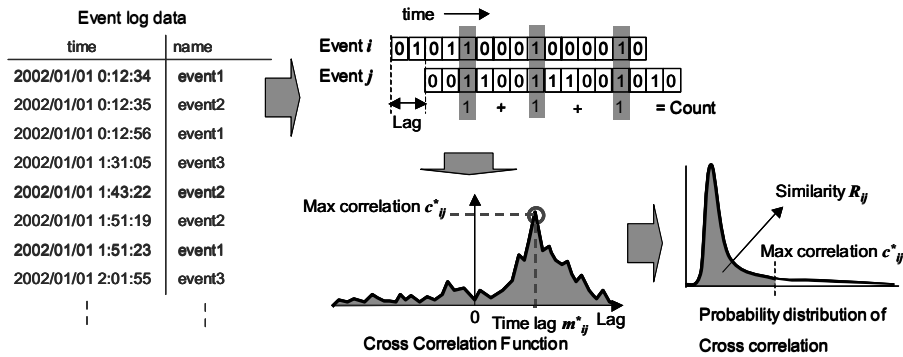
Figure 2. Event Correlation Analysis

2.3. Numerical Experiment

We evaluated the Event Correlation Analysis with synthetic data consists with six event type generated with the following rules for 30 days in one-second units.

- Event 1 and 4 are generated by Poisson distribution with mean of 0.0001 and 0.0004, respectively.
- Event 2 and 3 are generated from event 1, added with noise of Poisson distribution with mean of 0.000005 and removed with the probability 50%. In addition, occurrence time of each event is shifted with Normal distribution $N(600,100)$ and $N(1800,100)$, respectively.
- Event 5 and 6 are generated from event 2, added with noise of Poisson distribution with mean of 0.0002 and removed with the probability 50%. In addition, occurrence time of each event is shifted with Normal distribution $N(600,100)$ and $N(1800,100)$, respectively.

Table 3. Results for the synthetic data

Order $i \rightarrow j$	Similarity c^*_{ij}	Delay m^*_{ij}	Order $i \rightarrow j$	Similarity c^*_{ij}	Delay m^*_{ij}
1 –> 2	0.99	600	4 –> 5	0.98	600
1 –> 3	0.98	1800	4 –> 6	0.97	1900
2 –> 3	0.96	1200	5 –> 6	0.95	1100

Table 3 shows the similarity and delay time of the related event pairs with higher similarity than 0.9. From the table, we can see the proposed method properly extracted the related event pairs and their occurrence time lags even though the data contained noise.

3. Validation with Actual Plant Data

3.1. Target Plant

The proposed method was applied for actual chemical plant data for validation. The target plant has the typical multiple-stage gas purification unit. According to the event log data obtained from this unit, a total of 1,267 types of alarms and operator actions occurred a total of 56,350 times over a period of two months.

3.2. Results

The event log data was analyzed with Alarm Analyst R20, the software product that implements the Event Correlation Analysis. The software generated the groups with the consequential events for a few seconds. From the discussions with the plant engineers and operators, we were able to plan the solutions for IPL2&3 performance improvement regarding the top 35 groups accounted for 60% of occurrences within only six hours. We will next give some examples of the results, which show actual process property and operators' behavior correctly.

I. Consequential alarms

The five events listed in Figure 4 were occurred synchronously and the total count was 702. From the result of the Event Correlation Analysis, we found Alarm 1, 2, 3 and 4 followed after Operation 1 at around midnight everyday. From the interview with the operators, when they switched the blower to the manual operation, Alarm 4 notified them of pipe blowing. As a result, the air and steam flow rate became unstable and consequently Alarm 1, 2 and 3 occurred. Since these alarms did not have corresponding operator actions in this situation, it was decided that Alarm 1, 2 and 3 were inactive with alarm filtering technique during manual pipe blowing procedures. In addition, Alarm 4, which gave only guidance to the operators, was changed to a message.

Event	Description	1 Month
Alarm1	Upper and lower limit of tower air flow rate	+++ ++++++ ++++++++++++++++
Alarm2	Abnormality in steam flow rate	+ +++ +++++++++++++++++++++++++++++
Alarm3	Lower limit of air flow rate	+ +++ +++++++++++++++++++++++++
Alarm4	Message about blowing	+ +++ +++++++++++++++++++++++
Operation1	Blower auto start OFF	+ +++ +++++++++++++++++++++++
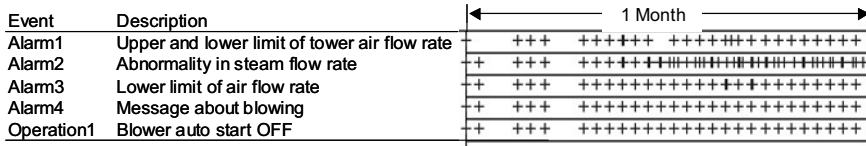
Figure 4. Example of consequential alarms

II. Complex operator actions

The five events listed in Figure 5 were also occurred synchronously and the total count was 495. From the result of the Event Correlation Analysis, we found Operation 2, 3 and Alarm 6 occurred simultaneously at around 4:00am, and Operation 2, 4 and Alarm 5 occurred simultaneously at around 4:00pm. The investigation of the daily report found that these events are related to the pipe flushing. In the system A, operators firstly invalidated alarms with Operation 2, followed by flushing procedure with Operation 3 and the notify of pipe flushing with Alarm 6. Likewise, in the system B, operators firstly invalidated alarms with Operation 2, followed by flushing procedure with Operation 4 and the notify of pipe flushing with Alarm 5. In general, operators should not intervene with such regular procedures. Therefore, Operation 2, 3 and 4 were automated with the timer, and Alarm 5 and 6 were changed to message.

Event	Description	1 Month
Operation2	Invalidating alarms for flushing system A and B	+ +++++++ +++++ ++ ++++++++++ ++ +++++++ ++ +++
Operation3	Flushing pipes for system A	++++++ +++++ ++++++ +++++++ ++++
Alarm5	Message indicating system B is being flushed	++++++ +++++ ++++++ +++++++ ++++
Operation4	Flushing pipes for system B	+++++++ +++++ ++++++ + +++++ +++
Alarm6	Message indicating system A is being flushed	+++++++ +++++ ++++++ + +++++ +++
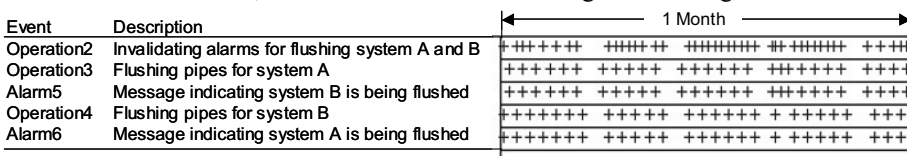
Figure 5. Example of complex operator actions

III. Redundant alarms

Alarm 7 is a flow rate alarm with lower limit type and occurred 657 times. According to the result of the Event Correlation Analysis, there were not any alarms or operator actions related to Alarm 7. Thus, this alarm is considered to occur independently with other alarms and operator actions. In fact, the operators told they usually did not take any actions in response to this alarm, but waited until it returned to stable by itself. We removed the alarms like this alarm that did not function as alarm, because they could deteriorate the sensitivity of the operators to alarms.

IV. Cause of upset

Figure 6 shows the example of the cause of plant upset. As shown in Figure 6, these six events occurred at the devices located near each other. From the event log data, Operation 5 was a manipulated variable change action, Operation 6 was an upper threshold change action, and Alarms 8, 9, 10 and 11 were all upper limit alarms. The occurrence order estimation found that Alarm 8 occurred first and Operation 6 occurred last. According to these results, it must be a problem with the pipes near the heat exchanger subjected to cooling with seawater. In fact, a field investigation revealed that the pipe easily became clogged, resulting in poor seawater cooling. Therefore, it was decided that this pipe should be cleaned regularly to solve this problem.
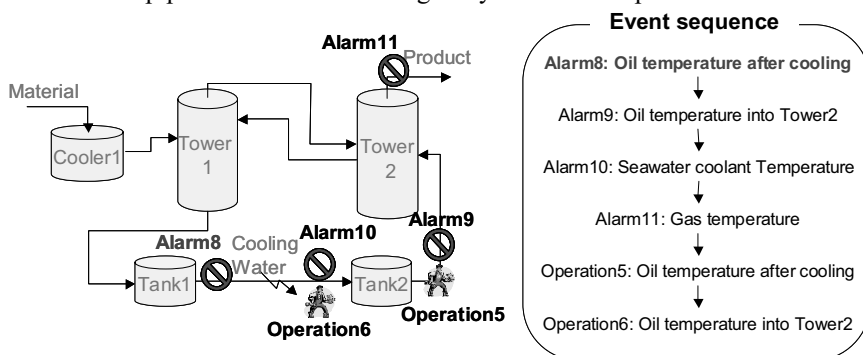
Figure 6. Example of cause of upset

4. Conclusion

In this paper, a new method for IPL2&3 performance improvement using Event Correlation Analysis was proposed. The validation result with the actual plant data shows our method is considered to improve process safety as well as productivity rapidly, easily, and effectively. In addition, since this method extracts the operators' behavior from the log data, it leads to standardization of operator actions.

References

AIChE/CCPS, Guidelines for Engineering Design for Process Safety, AIChE, 1993

Engineering Equipment & Materials Users' Association, Alarm Systems. A Guide to Design, Management and Procurement, EEMUA Publication No. 191, 1999

J. Shiozaki, B. Shibata, H. Matsuyama, and E. O'Shima, Fault Diagnosis of Chemical Processes Utilizing Signed Directed Graphs—Improvement by Using temporal Information, IEEE Transactions on Industrial Electronics, vol. 36, no. 4, pp. 469–474, 1989.

J. Nishiguchi and H. Tsutsui, 2005, A New Approach to Process Alarm Reduction Using Statistical Point Processes, SICE Annual conference 2005, pp.443-448

W. Li, Mutual Information Versus Correlation Functions, Journal of Statistical Physics, vol. 60, pp. 823-831, 1990

1227

Collaborative Multi - Agent based Process Monitoring System for Offshore Oil and Gas Production

Sathish S Natarajan, Kaushik Ghosh, Rajagopalan Srinivasan*

*Department of Chemical and Biomolecular Engg, National University of Singapore
10 Kent Ridge Crescent, Singapore, 119260.*

Abstract

Offshore oil and gas production platforms are uniquely hazardous in that operating personnel have to work in a perilous environment surrounded by extremely flammable hydrocarbons. A failure in an equipment could quickly propagate to others resulting in leaks, fires and explosions, causing loss of life, capital invested and production downtime. A method for preventing such accidents is to deploy intelligent monitoring tools which continuously supervise the process and the health of equipments to provide context–specific decision support to operators during safety-critical situations. A dynamic model of an offshore oil and gas production platform was developed using gPROMS and data to reflect operating conditions under normal, fault conditions and maintenance activities were simulated. These data are used to train three monitoring algorithms based on multivariate statistics (Principal Component Analysis), two of which are specialized in monitoring certain sections of the platform. These multivariate monitoring algorithms are considered as individual agents and the results produced by each are then integrated using a multi-agent collaborative framework. A consolidator agent, which uses voting based, Bayesian probability and Dempster Shafer fusion strategies for conflict resolution and decision fusion is developed. The ability of this agent based monitoring scheme to detect and diagnose faults in a more precise manner than any single FDI agent in offshore oil and gas production platforms is demonstrated.

Keywords: offshore production, process monitoring, agent based.

1. Introduction

Offshore Oil and Gas production has accounted for most of the increase in worldwide production of hydrocarbons over the last few decades. Given that most of the significant hydrocarbon discoveries in recent times have happened offshore, it is expected that the contribution of offshore production to our overall energy mix would only increase further. Hence it is imperative that these offshore production platforms are operated in a safe and efficient manner.

Operations personnel in offshore platforms face many unique challenges: The workforce lives and works in a restricted location for significant periods of time without a break. The environment is characterized by confined space, constant noise and activity while the production process has the potential for the rapid escalation of hydrocarbon related incidents. This inherently hazardous nature combined with the extremely harsh

*Corresponding Author
 Ph: +65 65168041, Email: chergs@nus.edu.sg

and stressful environment is the prime reason for the unusually large number of accidents. A study by the UK Health and Safety Executive shows that process related failures account for nearly 80% of the risk to offshore personnel. Currently, onboard operating personnel are expected to detect failures as and when they occur and take corrective action to prevent them from escalating into severe accidents. However given the cognitive limitations of operators working in a remote facility some failures are not detected.

A method for preventing accidents on offshore platforms is to deploy intelligent monitoring tools which continuously supervise the process and the health of the equipments to provide context – specific decision support to operators during safety critical situations. Historically, the offshore processing industry has sought to improve its safety performance using the proactive strategy in which the basic process design is made inherently safe and robust. The main shortcoming of this strategy is that even if all the hazards have been identified, it may not be economically or technically feasible to eliminate them. The reactive strategy seeks to overcome this through improved process control using advanced process monitoring and supervision. There is plenty of literature on process monitoring, fault detection and identification (Venkatasubramaniam et al, 2003) for petrochemical industries but these monolithic monitoring strategies are not suited for the offshore production process due to the lack of large scale computational facilities on board. Hence, a remote agent based distributed monitoring scheme would be more suited. Also, due to the complexity involved a single FDI agent would not be capable of handling all the fault scenarios. A method in which the results of several FDI agents, each specialized in monitoring certain sections of the overall process, are meaningfully fused by a consolidator agent, is proposed.

Since relevant process data for offshore platforms is not readily available in literature a dynamic model in gPROMS is developed, as detailed in Section 2. Process faults common in offshore production platforms (Khan et al, 2002) is introduced into this model. The operations data collected from this model was used to train three different multi-variate statistical models, each specialized in identifying a fault in a particular section of the platform, as discussed in Section 3. The results from these individual methods are then combined using different fusion strategies.

2. Dynamic Model of Offshore Natural Gas Production

The first step towards achieving a representative dynamic model is always a good steady state simulation. Hence a steady state model of a typical natural gas production platform was developed in Hysys. The steady state operations data from the Hysys model served as the initial conditions to the dynamic model in gPROMS. Three flow lines involving the usual combination of safety, master, wing and choke valves were modeled. The flow lines were combined into a production header and a separator header. As shown in Figure 1, a test separator is included which is brought online occasionally when a stream has to be analyzed. An important maintenance activity in the operation of these platforms is the pig-launching operation, during which the platforms are required to be shut down; hence shutdown and startup procedures were incorporated. Other maintenance activities include corrosion inhibitor and kinetic hydrate inhibitor injection. Apart from capturing the dynamics during these normal operations, three fault cases were simulated as indicated in Figure 1.
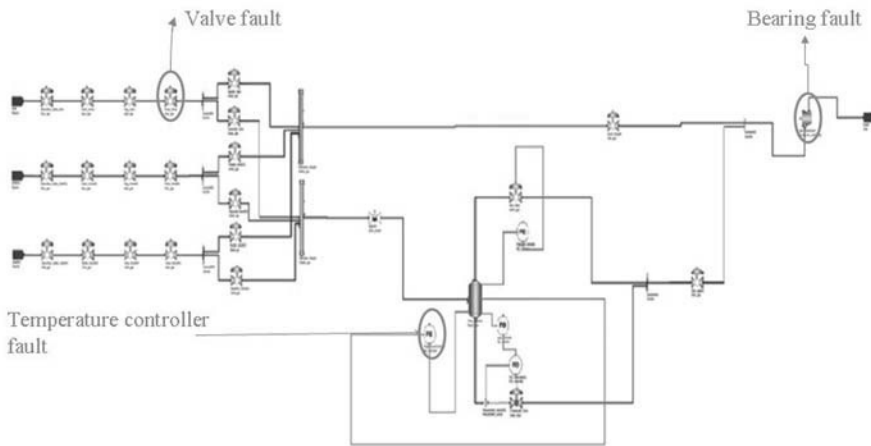
Figure 1. Process Flowsheet of Offshore Natural Gas Production from industrial case study

(1) A valve leak fault in a flow line in which the flow through the line was reduced from its normal range by 10% as a step change.

(2) Test separator temperature controller fault in which during the operation of the separator the heat input was reduced resulting in a temperature drop by 2%.

(3) A bearing fault on a compressor in which the eccentric vibrations and currents were simulated. This fault was not simulated as a complete failure of the bearing, but as an incipient fault like a dent in the raceways producing eccentricities.

Figure 2 shows the simulated data for these three faults along with their corresponding expected normal behavior.

3. Multivariate statistical models for monitoring

Three different models were developed as individual agents. The three agents were a process monitoring agent, which monitors the overall process, a specialized separator monitoring agent and a specialized compressor monitoring agent. The first two were based on principal component analysis (Jackson, 1991) while the third is based on spectral principal component analysis.

3.1. Process Monitoring Agent

The normal operating data from 34 variables measured at various points in the process were used to train a PCA model. The data from the three fault conditions were then projected onto this PCA model. The valve leak fault is detected and diagnosed with a delay of 14 samples. The temperature controller fault is detected with a delay of 4 samples and diagnosed after 16 samples. This process monitoring agent does not detect the bearing fault on the compressor.

3.2. Separator Monitoring Agent

Of the 34 variables measured, 5 were chosen around the separator. A separate PCA model is developed using normal data for these 5 variables and the fault data projected onto this. Being specially suited to detect abnormalities around the separator this agent detects and diagnoses the temperature controller fault with a delay of only 4 samples and is unable to detect the bearing fault around the compressor.
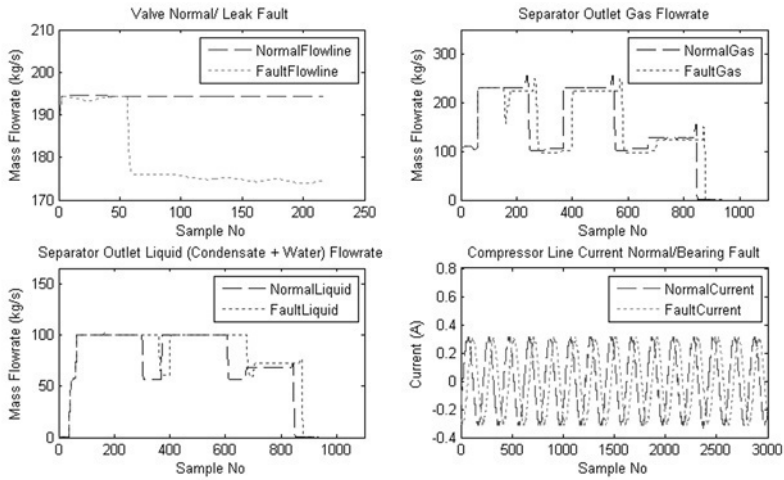
Figure 2. Simulated data for normal operating condition as well as Valve Leak Fault, Separator Temperature Controller Fault and Compressor Bearing Fault

3.3. Compressor Monitoring Agent

The eccentricities produced by the bearing fault are oscillatory in nature. Hence a Spectral PCA method is used for monitoring the compressor. The novel feature of spectral PCA is that the rows of the data matrix X, are the single sided power spectra $P(f)$ of the signals over a range of frequencies (Thornhill et al, 2002). The data matrix X is then decomposed in the usual PCA procedure and the T^2 and Q statistic measured. 5 variables were chosen around the compressor including the simulated vibration and current signals. This agent is capable of detecting and diagnosing the bearing fault alone after 11 samples.

4. Collaborative Multi-agent based fusion

As is evident from Table 1, no single FDI method is capable of detecting all the faults. Instead of developing a suitable monolithic method, the results from these individual methods could be meaningfully combined using a consolidator agent to produce good results. This consolidator agent utilizes voting, Bayesian, or Dempster Shafer fusion strategies. Using such methods, the strengths of the individuals agents are accrued while their drawbacks are overcome resulting in improved performance.
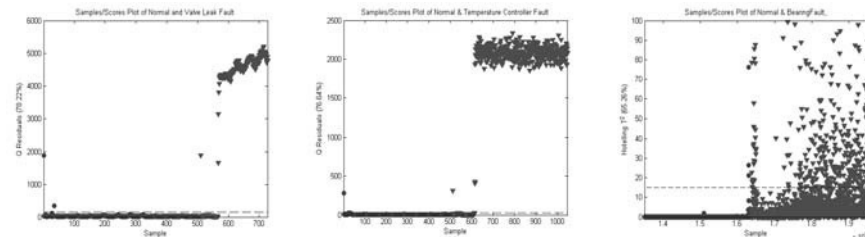
Figure 3. Sample/Score plot of Normal vs. Valve Leak Fault, Separator Temperature Controller Fault and Compressor Bearing Fault

Table 1. Performance of the individual agents, F1 - Valve Leak Fault, F2 - Separator Temperature Controller Fault, F3 – Compressor Bearing Fault

Fault	Process Monitoring PCA		Separator Monitoring PCA		Compressor Monitoring Spectral PCA	
	Detection Delay (x30s)	Diagnosis Delay (x30s)	Detection Delay (x30s)	Diagnosis Delay (x30s)	Detection Delay (x30s)	Diagnosis Delay (x30s)
F1	14	14	22	-	-	-
F2	4	16	4	4	-	-
F3	-	-	-	-	11	11
Average Delay	9	15	13	4	11	11
Recognition rate (%)	62.23		50.39		47.42	

Voting based fusion is the simplest and most commonly used decision fusion strategy. The class assigned to each sample by an FDI agent is considered as a vote for that class. The class with the maximum number of votes is the combined prediction. The major drawback of this method is that all agents are treated equally although in practice some agents outperform others and should hence be given more weightage.

Both Bayesian and Dempster-Shafer strategies utilize the prior performance of an agent for combination. The prior performance of a FDI agent is recorded in the form of a confusion matrix in which the rows are the actual classes while the columns are the assigned classes. For each FDI method, $P(C_i/C_j)$ is the probability of assigning a sample to class C_i given class C_j. For k FDI methods, Bayesian Belief Value for class C_i is calculated as

$$Bel(C_i) = \frac{\prod_k (P_k(C_i/C_j))}{\sum_i \prod_k (P_k(C_i/C_j))}$$

The class with maximum Bayesian belief value, *Bel(Ci),* is the assigned class upon combination using Bayesian fusion strategy.

Table 2. Performance of different fusion strategies, F1 – Valve Leak Fault, F2 – Separator Temperature Controller Fault, F3 – Compressor Bearing Fault

Fault	Voting Based Fusion		Bayesian Probability based Fusion		Dempster Shafer based fusion	
	Detection Delay (x30s)	Diagnosis Delay (x30s)	Detection Delay (x30s)	Diagnosis Delay (x30s)	Detection Delay (x30s)	Diagnosis Delay (x30s)
F1	22	-	14	14	14	22
F2	4	16	4	4	4	4
F3	-	-	11	11	11	11
Average Delay	13	16	9.67	9.67	9.67	12.33
Recognition rate (%)	49.53		97.85		97.17	

The Dempster-Shafer (D-S) method is popular in uncertainty reasoning which is used to combine separate pieces of evidence by assigning a belief value. The difference between Bayesian and D-S is the fact that the Bayesian method precisely assigns a probability value to each class, whereas D-S method assigns a BPA (Basic Probability Assignment) value to the union of all possible classes. The combined BPA values are obtained by using D-S combination rule.

$$Com\_BPA\left(C_i\right) = \frac{\displaystyle\sum_{A\cap B = C_i} BPA_I\left(A\right)BPA_{II}\left(B\right)}{1 - \displaystyle\sum_{A\cap B = \phi} BPA_I\left(A\right)BPA_{II}\left(B\right)}$$

Where Ci is the subset formed by the intersection of A and B and ϕ is null or empty set. The subset with highest combined BPA value is the resulting class upon combination.

The individual agents have recognition rates, which is defined as the percentage of sample correctly classified in the range of 47-62% shown in Table 1. Upon combination using Bayesian and Dempster Shafer based fusion strategies the recognition rates are improved to around 97% shown in Table 2. Voting based scheme does not perform as well as it requires at least 2 of the agents to concur for collaboration. We are able to achieve complete fault coverage and also improvement in fault detection and diagnosis time by using Bayesian and Dempster Shafer combination strategies.

5. Conclusion

An intelligent collaborative agent based monitoring system was proposed in this paper. Using data simulated from a dynamic gPROMS model three monitoring agents were developed, two of which were specialized. The results from these individual agents were then combined using different strategies. Upon combination we achieve significant (from about 60% to 97%) improvement in fault recognition rates and reduction in fault detection and diagnosis time. Thus, we have demonstrated that by using fusion we are able to combine the advantages of the individual methods while their respective shortcomings are overcome resulting in improved monitoring performance for offshore oil and gas production. Our future work would be to develop a remote agent based monitoring strategy, building upon these results for the offshore oil and gas industry, the first step of which would be the development of an ontological framework for efficient agent interaction.

Acknowledgement

References
1. Health and Safety Executive (HSE), 1996. Offshore accident/incident statistics reports. OTO96.954, London, UK.
2. V. Venkatasubramanian., R. Rengaswamy., S.N. Kavuri. 2003, A review of process fault detection and diagnosis , Computers and Chemical Engineering, Vol 27,3.
3. N.F.Thornill, S.L, Shah, B.Huang, A.Vishnubotla, 2002. Spectral principal component analysis of dynamic process data. Control engineering Practice 10, 833-846.
4. F.I. Khan., R. Sadiq., T. Husain., 2002, Risk-Based process safety assessment and control measures design for offshore process facilities., Journal of Hazardous Materials A94, 1-36.
5. J.E. Jackson., 1991, A User's Guide to Principal Components, Wiley, New York.

1233

Real-Time Moving Horizon State and Parameter Estimation for SMB Processes

Achim Küpper[a], Moritz Diehl[c], Johannes P. Schlöderl[b], Hans G. Bock[b],
Sebastian Engell[a]

[a]*Process Dynamics and Operations Group, Technische Universität Dortmund, Emil-Figge-Str. 70, 44221 Dortmund, Germany*
 (e-mail: achim.kuepper@bci.tu-dortmund.de)
[b]*IWR - Interdisciplinary Center for Scientific Computing, Universität Heidelberg, Germany*
[c]*Electrical Engineering Department (ESAT-SCD), K.U. Leuven, Belgium*

Abstract

Advances in numerical algorithms have rendered the application of advanced process control schemes feasible for complex chemical processes that are described by high-order first-principles models. Applying real-time iteration schemes reduces the CPU requirement such that rigorous models can be applied that enable a precise forecast of the system behaviour. In this paper, a moving horizon state and parameter estimation scheme for chromatographic simulated moving bed SMB processes is presented. The simultaneous state and parameter estimation is based on a high-order nonlinear SMB model which incorporates rigorous models of the chromatographic columns and the discrete shifting of the inlet and outlet ports. The estimation is performed using sparse measurement information: the concentrations of the components are only measured at the two outlet ports (which are periodically switched) and at one fixed location between two columns. The goal is to reconstruct the full state of the system, i.e. the concentration profiles along all columns, and to identify model parameters reliably. The state estimation scheme assumes a deterministic model within the prediction horizon, state noise is only present in the state and in the parameters prior to and at the beginning of the horizon. The scheme can be applied online. The advantage of this estimation scheme is that it is applicable to all process scenarios encountered during the real operation of an SMB plant, e.g. start up, transition periods, varying flows and switching times, since no model simplification nor a state reduction scheme are applied. Numerical simulations (start up of the SMP process) of a validated model for a separation problem with nonlinear isotherms of the Langmuir type demonstrate the efficiency of the algorithm.

Keywords: Simulated Moving Bed chromatography, Moving horizon estimation, State estimation, Model identification, Real-time application, Real-time iteration.

1. Introduction

The Simulated Moving Bed (SMB) process is an efficient chromatographic separation technology that is increasingly applied in the food, fine chemicals, and pharmaceutical industries. A SMB process is realized by connecting several chromatographic columns in a closed loop as illustrated by Fig. 1. Since SMB processes and its variants are characterized by mixed discrete and continuous dynamics, spatially distributed state vari-

ables with steep slopes, and slow and strongly nonlinear responses of the concentrations profiles to changes of the operating parameters, they are difficult to control. In the literature, relatively few contributions that deal with state estimation of SMB processes can be found. The published work is either based upon simplifications or deals with the engineering of tailored estimation schemes. Recently, a rigorous moving horizon estimation approach for SMB processes was proposed in [2]. In this formulation of the MHE, a deterministic behaviour of the process on the estimation horizon and Gaussian independent identically distributed measurement noise are assumed. The initial state at the beginning of the horizon and its covariance are computed by an Extended Kalman Filter (EKF). The state noise covariance and the initial error covariance of this EKF are the only tuning parameters of this scheme. A fast online solution of the underlying constrained least-squares optimization problem is obtained by using the direct multiple shooting method [3]. A full rigorous model is applied and therefore no assumption that the plant is close to the periodic steady state is needed. Along with the states, key adsorption parameters are estimated online. Simulations demonstrate that the states and critical model parameters can be reconstructed successfully. The computation times are such that the estimator can be applied online. Since a rigorous full scale SMB model is used, the MHE approach of [2] can also handle transition periods. In this contribution, the moving horizon state and parameter estimation scheme is applied to reconstruct the state variables in the start-up phase.

2. The Simulated Moving Bed Process

Chromatographic separation is based on the different adsorption affinities of the molecules in the liquid to an adsorbent which is packed in a chromatographic solid bed. A counter-current movement between the liquid and the adsorbent is implemented by switching the ports in the direction of the liquid flow periodically, illustrated by Fig. 1.
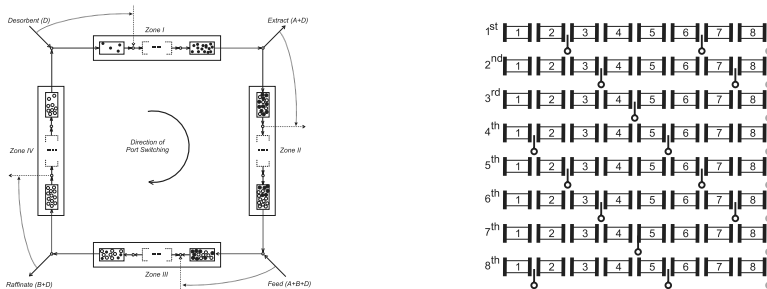
Fig. 1: Principle of the Simulated Moving Bed process (left) and positions of the extract, raffinate, and recycle port measurements over a cycle of operation (right)

In this paper, the counter-current flow of the solid and the liquid phases is modeled as it is achieved in the real plant by periodically switching the inlet and outlet ports in the direction of the liquid flow after period m with period length τ has passed. The state variables thus represent the concentrations in the physical columns and do not exhibit jumps. Only the input flow rates and the inflow concentrations change discontinuously. The dynamic simulation of the SMB process is achieved by integrating the differential equation over period m

$$\dot{x} = f(x, Q_m, p) \quad t \in [(m-1)\tau, \tau], \qquad x(t_0) = x_{m,0} \tag{1}$$

followed by the switching of the flows $Q_{m,j}$:

$$Q_{m+1,j} = M_Q Q_{m,j} \qquad j = De, Ex, Fe, Ra, \mathrm{Re} \tag{2}$$

with differential states x and parameters p. The vector $Q_{m,j}$ defines the inlet/outlet flow of port j (desorbent, extract, feed, raffinate) and the recycle stream at period m. The respective component of $Q_{m,j}$ represents the flow of the respective port j to the respective column. M_Q is a permutation matrix that shifts the flow ports after period m has passed. The recycle flow that defines the total flow rate in the zone in front of the desorbent port is also switched in the same manner. All zone flow rates result from the port flows and the recycle flow. For this paper, three positions where the concentrations of the two substances of the mixture are measured are assumed. The measurements are positioned behind the extract port, behind the raffinate port, and behind one column in the process where physically the closing of the loop is realized. The extract and raffinate concentration measurements move together with the ports while the third measurement remains at its physical location throughout. Fig. 1 illustrates the measurement positions over a full cycle of operation that is defined by the time that passes until the ports are back at the starting positions after having moved around the closed loop of columns. The positions of the inlet and outlet ports are then repeated periodically. From mass balances of the components around the inlet and the outlet ports, the internal flow rates and the inlet concentrations can be calculated according to:

$$\begin{aligned}
\text{Desorbent node}: \quad & Q_{IV} + Q_{De} = Q_I \,; c_{i,IV}^{out} Q_{IV} = c_{i,I}^{in} Q_I \quad i = A, B \\
\text{Extract node}: \quad & Q_I - Q_{Ex} = Q_{II} \\
\text{Feed node}: \quad & Q_{II} + Q_{Fe} = Q_{III} \,; c_{i,II}^{out} Q_{II} + c_{i,Fe} Q_{Fe} = c_{i,III}^{in} Q_{III} \quad i = A, B \\
\text{Raffinate node}: \quad & Q_{Ra} + Q_{IV} = Q_{III}
\end{aligned} \qquad (3)$$

where $Q_{I\text{-}IV}$ are the flow rates in the corresponding zones, Q_{De}, Q_{Ex}, Q_{Fe}, and Q_{Ra} denote the external flow rates and $c_{i,in}$ and $c_{i,out}$ denote the concentrations of the component i in the streams leaving and entering the respective zone. The column distribution over the four separation zones considered here is 2/2/2/2, i.e. each zone consists of two columns. The chromatographic columns are modelled by the General Rate Model where it is assumed that there are no radial gradients in the column and that the particles of the solid phase are uniform, spherical, porous (with constant particle porosity ε_p), and that the mass transfer between the particle and the surrounding layer of the bulk is in a local equilibrium. The concentration of component i is denoted by c_i in the liquid phase and by q_i in the solid phase. D_{ax} is the axial dispersion coefficient, u the interstitial velocity, ε_p the void fraction of the bulk phase, $k_{l,l}$ 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 partial differential equations of a column can be derived from a mass balance around an infinitely small cross section area of the column with a constant radial distribution of the interstitial velocity u and the concentration c_i.

$$\frac{(1-\varepsilon_b)3k_{l,i}}{\varepsilon_b}(c_{b,i} - c_{b,i|r=R_p}) = D_{ax}\frac{\partial^2 c_{b,i}}{\partial^2 z} - u\frac{\partial c_{b,i}}{\partial z}; \quad (1-\varepsilon_p)\frac{\partial q_i}{\partial t} + \varepsilon_p\frac{\partial c_{p,i}}{\partial t} = \left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial c_{p,i}}{\partial r}\right)\right], \quad (4)$$

with appropriate initial and boundary conditions. It is assumed that the concentration q_i is in thermodynamic equilibrium with the liquid concentrations in the particle and their relationship can be described by an adsorption isotherm. In the application considered here, the adsorption isotherm is of extended Langmuir type

$$q_i = H_i^1 c_{p,i} + \frac{H_i^2 c_{p,i}}{1 + \sum_k k_k c_{p,k}}, \quad i = A, B, \qquad (5) \quad \text{with } H_i^j \text{ and } k_i \text{ as isotherm constants.}$$

The resulting system of coupled differential equations can be efficiently solved by the numerical approach proposed in [4] where a Galerkin finite element discretization of the bulk phase is combined with an orthogonal collocation of the solid phase. As a result,

the initial values, boundary values, and partial differential equations (PDE) are transformed into a set of initial values and a system of ordinary differential equations (ODE) $\dot{x} = f(x, u, p)$ of order 352 (for n_{fe}=5, n_c=1, number of components n_{sp}=2, and number of columns n_{col}=8) where the flows Q are summarized in the input vector u. The system output is defined as $y=h(x,u,p)$.

3. Moving Horizon Estimation in SMB Processes

For the simultaneous estimation of the states and the parameters of SMB processes, we employ the Moving Horizon Estimation scheme introduced by Diehl, Schlöder, and Bock [2], which is modified in order to handle the shift of the inputs and of the measurements of the SMB process. The Moving Horizon Estimator estimates the states and the parameters based on the past measurements at specific time points that are located in the horizon $T_N=t_K-t_L$. t_K represents the current time and t_L is the time at the beginning of the horizon. A least-squares minimization is performed that minimizes the deviations of the real measurements η_k from the simulated measurements $h(x;t_k,u,p)$ at times t_k. The optimization problem of the MHE results as:

$$\min_{x(t_L),p} \left\{ \left\| \begin{matrix} x(t_L) - \bar{x}_L \\ p - \bar{p}_L \end{matrix} \right\|_{P_L^{-\frac{1}{2}}}^2 + \sum_{k=L}^{K} \left\| \eta_k - h(x(t_k),u(t_k),p) \right\|_{V_k^{-\frac{1}{2}}}^2 \right\} \quad \begin{matrix} s.t. \quad \dot{x}(t) = f(x(t),u(t),p) \quad (6) \\ x_{\min} \le x \le x_{\max}; \quad p_{\min} \le p \le p_{\max}; \quad t \in [t_L, t_K] \end{matrix}$$

The second term represents the prediction errors within the horizon and the first term represents the arrival cost (the penalization of a change of the estimates of the initial values of the states and of the parameters). Note that only the initial values of the states and the parameters are free parameters of the optimization problem because no state noise is assumed within the horizon, which is equivalent to the assumption of a dynamic model that is perfect in its structure. The absence of state noise on the horizon is compensated by the simultaneous estimation of key model parameters which is an appropriate assumption since uncertainties are mostly due to model errors and not to disturbances. From the solution $x(t_L)$ and p of the optimization problem, the deterministic model is simulated forward to obtain the current estimated state x_K. The MHE takes upper and lower bounds on the states and on the parameters into account. The means $\bar{x}(t_L)$ and $\bar{p}$ and the covariance P_L in the arrival cost are determined by a smoothed Extended Kalman Filter that is implemented using QR-decompositions in order to guarantee positive definite matrices despite possible numerical errors [5]. In order to enable a smooth gradient calculation with respect to the simulated measurements which are subject to jumps due to the periodic movement of the ports, virtual measurements at constant positions at the outlet of each chromatographic column are included in the mapping of h. In order to account for the actual existence of real measurements at the considered time point k, the corresponding components on the diagonal of the measurement weight V_k^{-1} are set to σ_v^{-1} while nonexisting measurements have zero entries on the diagonal of V_k^{-1}. A zero weight can be interpreted as infinite measurement noise. The last two entries of V_k are the variances of the measurements at the internal measurement position (recycle) which are not shifted.

3.1. Multiple-Shooting Real-Time Iteration Scheme for MHE

The moving horizon optimization problem is solved by the multiple shooting methods that subdivides the time horizon into subintervals and formulates autonomous initial value problems on each individual subinterval. To retain the original problem, the subproblems are coupled by continuity conditions to guarantee the continuity of the state trajectories. The computational requirements are largely reduced by applying the *real-*

time iteration scheme for the multiple shooting method introduced by [1] that updates all relevant sensitivity matrices that are necessary to solve the optimization problem before the most recent measurement η_K is available. Another important feature is that the next optimization problem is initialized well at the current solution such that the number of iterations can be reduced to one.

3.2. Results (start-up scenario)

The separation of the enantiomer mixture EMD-53986 into the pure components is considered. The parameters of the SMB model were taken from [6].The measurements are corrupted by noise with a standard deviation of *0.025 g/l*. No cross-correlations between the state noises and between the state noises and the parameters were assumed. The same noise variances were assumed for each state. The tuning of the moving horizon estimator was performed by varying the covariances of the state variables and of the free parameters. The weighting matrix *W* incorporates a state standard deviation of 0.00433 *g/l* and a parameter standard deviation of 0.0316 for H^2_A and 0.0265 for H^2_B: $W^{0.5}=diag(0.0043,...,0.0043,0.0361,0.0265)$. The initial weight P_0 is set to 0.005 *W*. The chosen state and parameter noises represent a compromise between the smooth estimation of the states and a quick adaptation of the parameters. The sampling time of the estimator is 1/10 of the period length. The moving horizon length is five sampling intervals (half a period). In the simulation scenario, the reference model and the model of the estimator are initialized with empty columns.The output, parameters, and states are reconstructed correctly, as can be seen by figures 2 to 4. The estimation of the parameters does not interfere with the reconstruction of the states. The MHE estimator can be applied online, as can be seen from Fig. 3. The CPU times are below the sampling rate at all sampling points. The CPU times of the MHE are around 19 s on a standard PC (Intel Xenon CPU 2.8 GHz, 4.0 GB RAM), the maximum and minimum values being 25.0 s
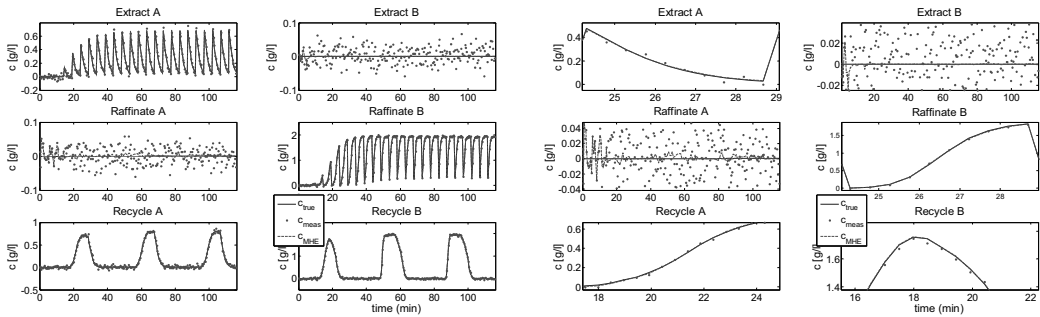
Fig. 2: Measurements (extract, raffinate, recycle); enlarged (right)
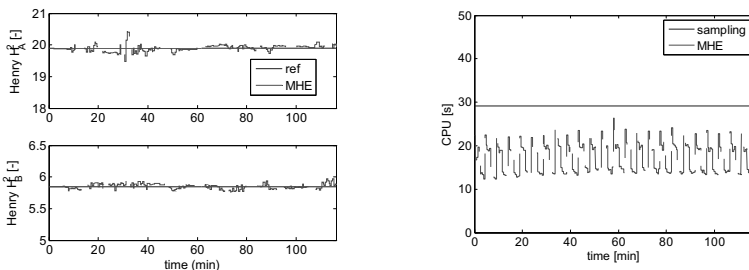
Fig. 3: Parameter estimates (left) and CPU times of the estimator at each sampling point (right)

t=0.0 min, (2,6,8)   t=0.97 min, (2,6,8)   t=4.37 min, (2,6,8)   t=6.32 min, (3,7,8)

t=9.23 min, (3,7,8)   t=14.09 min, (4,8)   t=18.95 min, (1,5,8)   t=38.39 min, (1,5,8)
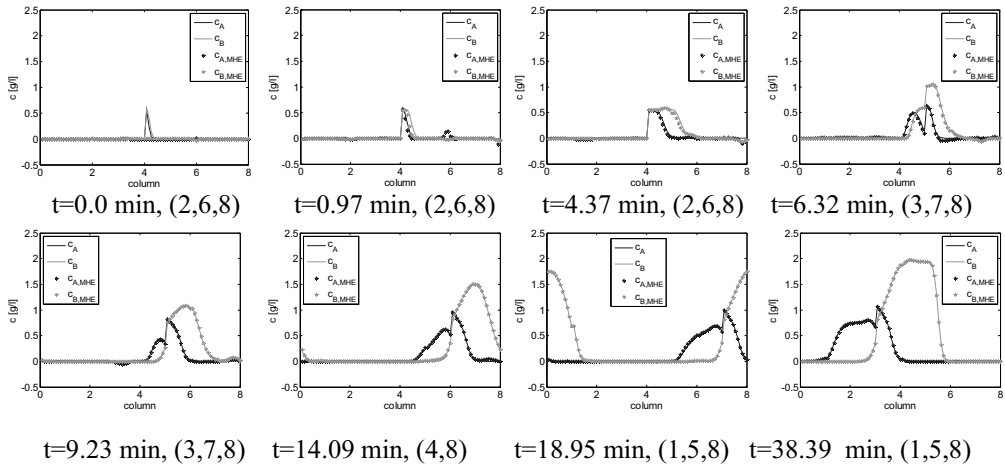
Fig. 4: Start-up scenario: real and estimated concentrations; measurements at the outlet of the columns indicated in brackets

4. Conclusion

In this paper, a moving horizon state estimation scheme that successfully reconstructs the states and key parameters was described and applied to a high-order nonlinear process with discrete switching of the inputs, the Simulated Moving Bed process. The integration of the presented moving horizon scheme in an advanced control scheme is subject of ongoing research.

5. Acknowledgements

References

[1] M. Diehl, H. Bock, J. Schlöder, R. Findeisen,Z. Nagy, F. Allgöwer (2002), Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations, *Journal of Process Control*,12(4), 577-585.

[2] A. Küpper and M. Diehl and J. Schlöder and H.G. Bock and S. Engell (2009), Efficient moving horizon state and parameter estimation for SMB processes, *Journal of Process Control*, doi:10.1016/j.jprocont.2008.10.004.

[3] H.G. Bock (1981), Numerical treatment of inverse problems in chemical reaction kinetics, in Modelling of Chemical Reaction Systems,K.H. Ebert and P. Deuflhard and W. Jäger (editors), 18 of *Springer Series in Chemical Reaction Systems*, Springer-Verlag.

[4] T. Gu (1995), Mathematical modelling and scale-up of liquid chromatography, Springer Verlag, New York.

[5] M. Diehl(2002), Real-time optimization of large scale nonlinear processes, 920 of *Fortschritt-Berichte VDI Reihe 8, Mess-, Steuerungs- und Regelungstechnik*, VDI Verlag, Düsseldorf.

[6] A. Jupke (2003), Experimentelle Modellvalidierung und modellbasierte Auslegung von Simulated Moving Bed (SMB) Chromatographieverfahren, Dr.-Ing. Dissertation, Fachbereich Bio- und Chemieingenieurwesen, Universität Dortmund, Shaker-Verlag, Aachen.

Channel Blockage Detection of Microreactors Using Pressure Sensors

Yasuyuki Kaburagi, Masaru Noda and Hirokazu Nishitani

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0192, Japan

Abstract

Blockage in microreactors is a serious problem that limits their practical usage. A proper diagnosis of blockage is indispensable to ensure these machines operate effectively and stably. We describe two systems, A and B, that can locate channel blockage using pressure sensors. System A uses a computational fluid dynamic (CFD) simulation, and system B uses a simple pressure balance model to construct databases for channel blockage diagnosis. The results of case studies showed that system A can identify the blockage location accurately using fewer pressure sensors than microchannels. System B can also identify the blockage location in many cases, but several misdiagnoses occur. In addition, we describe a method that estimates the degree of blockage in a microchannel. Simulation results showed that our method can estimate the degree of blockage precisely using the pressure sensor that is closest to the blocked location identified by systems A or B.

Keywords: Microreactor, Channel blockage diagnosis, Support vector machine, Computational fluid dynamics, Pressure balance model

1. Introduction

Microreactors have received both industrial and academic attention as a new type of production technology for specialty chemicals that are difficult to produce with conventional reactors. Blockage in the microchannels of microreactors is a serious problem that limits their practical usage. It changes the flow pattern in the microreactor and degrades the product quality. Detecting and identifying the blockage location is indispensable for ensuring more effective and stable operation in these microreactors.

Kano *et al.* (2007) proposed a blockage diagnosis system for a stacked microreactor that uses temperature sensors. The system uses the temperature differences at sensor locations between normal and abnormal operating conditions to identify the blockage location. When the channel blockage does not affect the temperature in the microchannels due to a high surface/volume ratio, which is one of the characteristics of microreactors, the system may not work. Also, they did not address how to estimate the degree of blockage after identifying the location.

This paper describes two blockage diagnosis systems that use pressure sensors instead of temperature sensors. One uses a computational fluid dynamic (CFD) simulation, and the other uses a simple pressure balance model for constructing a database for a channel blockage diagnosis. We also describe a method for estimating the degree of blockage after the location is identified. A CFD simulation was also used in this study to generate the validation data for our systems in case studies.

2. Microreactor

A microreactor generally consists of three parts, namely, an inlet manifold for flow distribution, parallelized microchannels for the reaction, and an outlet manifold for mixing, as shown in Fig. 1. The industrial scale production using a microreactor requires a high number of parallelized microchannels because each one provides only a small amount of product. The blockage in the microchannel causes poor uniformity in the residence time distribution between them, which may make product quality worse.

This study introduces the following assumptions for developing a blockage diagnosis system of a microreactor:

Fig. 1 Structure of microreactor

(1) Pressure sensors are set up at the inlets of microchannels. The total number of sensors, N_C, and their locations, $s \in S$, are given in advance.
(2) Channel blockage occurs in only one microchannel at a time.

3. Blockage diagnosis system A based on CFD simulation data

When channel blockage occurs in a microreactor, the pressure distribution changes. If the pressure distribution has certain patterns due to the blockage location, comparing measured pressure-distribution data with the prepared pressure distribution data when channel blockage occured can identify the blockage location.

In this study, we used a CFD simulation to execute a number of simulations for preparing pressure distribution data at sensor locations under various degrees of blockage in each microchannel. The pressure distribution data calculated by the CFD simulation are used as a training data set for a support vector machine (SVM) to construct a discriminate function for identifying the blockage location.

The detailed procedure for building System A is as follows. We assume that pressure distribution data vector $\widetilde{P}_N$ consisting of measured pressure $\widetilde{p}_{N,s}(s \in S)$ in a normal operating condition is known. Here, $\widetilde{p}_{N,s}$ denotes the pressure data measured at sensor location s without any blockage, and S is the set of sensor locations.

(1) Calculate pressure data vector $\hat{P}_{B,i,x}$ consisting of $\hat{p}_{B,i,x,s}(s \in S)$ for some degrees of blockage $x \in X$ in microchannel i using the CFD simulation. Here, $\hat{p}_{B,i,x,s}$ denotes the calculated pressure data at sensor location s under $x\%$ blockage in microchannel i ($i = 1,\ldots,N_C$). X is the set of degrees of blockage for which pressure data is calculated using the CFD simulation.
(2) Calculate the pressure difference vector $\Delta \hat{P}_{i,x}$ by Eq.(1).

$$\Delta \hat{P}_{i,x} = \hat{P}_{B,i,x} - \widetilde{P}_N \quad (i = 1,\ldots, N_C, x \in X) \tag{1}$$

(3) Construct a discriminate function using the SVM. The database of blocked channel i and pressure difference data $\Delta\hat{P}_{i,x}$ $(i = 1,..., N_C, x \in X)$ are used as a training data set for the SVM.

The blockage location is identified through the following procedure:
(1) Obtain actual pressure data vector $\widetilde{P}_B$ consisting of $\widetilde{p}_{B,s}$ $(s \in S)$.
(2) Calculate pressure difference data vector $\Delta\widetilde{P}$ using Eq.(2)

$$\Delta\widetilde{P} = \widetilde{P}_B - \widetilde{P}_N \tag{2}$$

(3) Identify the blockage location using the discriminate function of the SVM.

4. Blockage diagnosis system B based on pressure balance model

A CFD simulation is a powerful tool for estimating the states in a microreactor when channel blockage occurs, but it is a time-consuming and costly process. Commenge *et al.* (2002) proposed a pressure balance model (PBM). Figure 2 shows an example of compartments defined in the PBM. The PBM consists of a simplified description of pressure and material balances in compartments. They applied this model to estimate states such as the flow rate and pressure drop in each microchannel instead of using a CFD simulation.
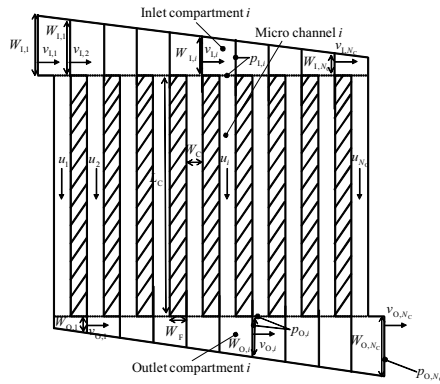
Fig. 2 Pressure balance model

This section describes a simple blockage diagnosis system, B, that is based on the pressure balance model. System B uses correlations between the pressure data vector calculated using the PBM under 50% blockage in each microchannel and the measured pressure data vector at the sensor locations. The procedure for building the database for system B is as follows:

(1) Calculate pressure data vector $\overline{P}_{B,i}$ consisting of $\overline{p}_{B,i,50,s}$ $(s \in S)$ under 50% blockage in microchannel i $(i = 1,…,N_C)$ using the PBM.
(2) Calculate pressure difference data vector $\Delta\overline{P}_i$ using Eq.(3).

$$\Delta\overline{P}_i = \overline{P}_{B,i} - \widetilde{P}_N \quad (i = 1,..., N_C) \tag{3}$$

The blockage location is identified through the following procedure:
(1) Obtain actual data vector $\widetilde{P}_B$ consisting of $\widetilde{p}_{B,s}$ $(s \in S)$.
(2) Calculate pressure difference data vector $\Delta\widetilde{P}$ using Eq.(2)

(3) Identify microchannel i^*, which has the highest correlation between $\Delta\widetilde{P}$ and $\Delta\overline{P}_i$ as the blocked microchannel. The correlation R_i is defined using Eq.(4).

$$R_i = \frac{(\Delta \widetilde{P} - \Delta \overline{\widetilde{P}})^T (\Delta \overline{P}_i - \Delta \overline{\overline{P}}_i)}{\left\| \Delta \widetilde{P} - \Delta \overline{\widetilde{P}} \right\| \left\| \Delta \overline{P}_i - \Delta \overline{\overline{P}}_i \right\|} \tag{4}$$

Here, all elements of $\Delta \overline{\widetilde{P}}$ and $\Delta \overline{\overline{P}}_i$ are the means of all elements of $\Delta \widetilde{P}$ and $\Delta \overline{P}_i$, respectively. The superscript T denotes the transpose, and $\left\| \cdot \right\|$ denotes 2-norm.

5. Estimation method of degree of blockage

After the location of the blockage is identified, the degree is estimated through the following procedure.

(1)   Obtain pressure data $\widetilde{p}_{B,i^{**}}$ at sensor location i^{**}, which is the closest location to microchannel i^{*} identified by systems A or B as the blocked location.

(2)   Determine $\overline{W}_C^{*}$, which minimizes the objective function J given using Eq.(5).

$$J = \left| \widetilde{p}_{B,i^{**}} - \overline{p}_{B,i^{**}} (W_C^{*}) \right| \tag{5}$$

Here, W_C^{*} is the width of microchannel i^{*}. $\overline{p}_{B,i^{**}}$ is the pressure data at sensor location i^{**} calculated using the PBM.

(4)   Determine the degree of blockage (BD) of microchannel i^{*} using Eq.(6).

$$BD = \left(1 - \frac{Z_C \cdot \overline{W}_C^{*} \cdot \overline{u}_{B,i^{*}}}{Z_C \cdot W_C \cdot \overline{u}_{N,i^{*}}} \right) \times 100 \ \ [\%] \tag{6}$$

Here, Z_C and W_C are the original depth and width of each microchannel. $\overline{u}_{B,i^{*}}$ and $\overline{u}_{N,i^{*}}$ denote the flow rate in microchannel i^{*} with and without blockage calculated using the PBM, respectively.

6. Case study

Systems A and B were applied to the blockage diagnosis problems of a microreactor with 10 microchannels. The geometric parameters in Fig. 2 and the operating conditions of the microreactor are summarized in Table 1. The following four types of pressure sensor configurations were implemented as different case studies. In this study, Fluent version 6.1^{TM} (ANSYS Japan) was used to generate validation data for our systems.

Configuration 1: N_S=10, S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
Configuration 2: N_S=5, S = {1, 3, 5, 7, 9}
Configuration 3: N_S=4, S = {1, 4, 7, 10}
Configuration 4: N_S=3, S = {1, 5, 9}

Eighty cases of CFD simulations were executed to obtain pressure distribution data under eight degrees of blockage: 4.8%, 9.1%, 18.5%, 36.4%, 44.8%, 52.9%, 70.8%, and 79.1% in each microchannel. The results of the CFD simulation were used as a training data set to construct a discriminate function of the SVM.

The diagnosis results of system A are summarized in Table 2. The identified number of the blocked microchannel is shown in Table 2 for ten blockage locations, for three

degrees of blockage (8.6%, 27.3%, and 61.4%), and for four sensor configurations. System A perfectly identified the blockage location using only four pressure sensors (Configuration C). Four misdiagnoses, including failure to diagnose, occurred when the degree of blockage was 8.6% and when sensor configuration D was used. These simulation results proved the usefulness of system A.

System B was also applied to the aforementioned blockage diagnosis problem. Table 2 shows that system B identified the blockage location in many cases, but several misdiagnoses occurred. A misdiagnosis tends to occur when the degree of blockage is low and when the pressure sensor interval is high. This is due to a mismatch of the PBM and the model used in the CFD simulation.

Table 3 shows the estimation results of the degree of blockage. The results show that our method can estimate the degree of blockage with less than 2.1% error.

Table 1 Geometric parameters and operation conditions of microreactor

Name	Parameter	Value	Unit	Name	Parameter	Value	Unit
Number of channel	N_C	10	-	Inlet manifold	$W_{I,1}$	5.0	mm
Channel width	W_C	100	μm		$W_{I,Nc}$	1.4	mm
Channel depth	Z_C	100	μm	Outlet manifold	$W_{O,1}$	1.4	mm
Channel length	L_C	20	mm		$W_{O,Nc}$	5.0	mm
Width of fin	W_F	284	μm	Inlet velocity	$v_{I,1}$	0.01	m/s
Viscosity	M	0.1	Pa·s	Outlet pressure	$p_{O,Nc}$	101.3	kPa

7. Conclusion

Two blockage diagnosis systems, A and B, that use pressure sensors were described. Simulation results showed that system A can diagnose the blockage location successfully using only four pressure sensors against ten microchannels, even when the degree of blockage is low. Though system B can also identify the location of blockage in many cases, several misdiagnoses occurred. The diagnosis performance of system B is worse than that of system A, but it will be useful when CFD simulation is not available.

When some experimental data under abnormal operating conditions are obtained, those data can be used as a part of a training data set for constructing a discriminate function of system A. An optimal integration of experimental data and CFD simulation data is a next problem to be discussed.

Our blockage diagnosis system and the estimation method for the degree of blockage are very useful for detecting early blockage and for identifying the blocked microchannels. In this study, we focused on a microreactor with many microchannels, but our methods can very easily be applied to various types of micro-devices when pressure distribution data can be calculated under abnormal operating conditions.

References

Commenge, J. M., L. Falk, J. P. Corriou and M. Matlosz, 2002, Optimal Design for Flow Uniformity in Microchannel Reactors, *AIChE J.*, 48, 2, 345-358

Kano, M., T. Fujioka, O. Tonomura, S. Hasebe and M. Noda, 2007, Data-Based and Model-Based Blockage Diagnosis for Stacked Microchemical Processes, *Chem. Eng. Sci.*, 62, 1073-1080

Table 2 Diagnosis results of systems A and B

Blockage location	Blockage degree [%]	System A				System B			
		A	B	C	D	A	B	C	D
	61.4	1	1	1	1	1	1	1	1
1	27.3	1	1	1	1	1	1	1	1
	8.6	1	1	1	1	1	1	1	1
	61.4	2	2	2	2	2	2	1	1
2	27.3	2	2	2	2	2	2	1	1
	8.6	2	2	2	1	2	2	1	1
	61.4	3	3	3	3	3	3	4	3
3	27.3	3	3	3	3	3	3	4	3
	8.6	3	3	3	2	3	3	3	4
	61.4	4	4	4	4	4	4	5	5
4	27.3	4	4	4	4	4	4	5	5
	8.6	4	4	4	4	4	4	5	5
	61.4	5	5	5	5	5	5	5	6
5	27.3	5	5	5	5	5	5	5	6
	8.6	5	5	5	5	5	5	5	6
	61.4	6	6	6	6	6	6	6	6
6	27.3	6	6	6	6	6	6	6	6
	8.6	6	6	6	6	6	6	7	6
	61.4	7	7	7	7	7	7	7	7
7	27.3	7	7	7	7	7	7	7	7
	8.6	7	7	7	7	7	7	7	7
	61.4	8	8	8	8	8	8	8	8
8	27.3	8	8	8	8	8	8	8	8
	8.6	8	8	8	-*	8	8	7	10
	61.4	9	9	9	9	9	9	9	9
9	27.3	9	9	9	9	9	9	9	9
	8.6	9	9	9	9	9	9	9	9
	61.4	10	10	10	10	10	9	10	9
10	27.3	10	10	10	10	10	9	10	9
	8.6	10	10	10	8	10	10	10	8

* Fail to identify

Table 3 Estimation results of degree of blockage (*BD*) of microchannel

Blockage location	Actual *BD* [%]	Estimated *BD* [%]	Blockage location	Actual *BD* [%]	Estimated *BD* [%]
	61.4	61.4		61.4	59.7
1	27.3	27.1	6	27.3	26.6
	8.6	8.4		8.6	8.3
	61.4	59.7		61.4	59.5
2	27.3	26.5	7	27.3	26.4
	8.6	8.3		8.6	8.3
	61.4	59.5		61.4	60.0
3	27.3	26.4	8	27.3	26.7
	8.6	8.3		8.6	8.4
	61.4	59.3		61.4	61.4
4	27.3	26.4	9	27.3	27.3
	8.6	8.3		8.6	8.4
	61.4	60.9		61.4	60.6
5	27.3	26.9	10	27.3	26.9
	8.6	8.6		8.6	8.4

PETROBRAS Experience Implementing Real Time Optimization

Fábio S. Liporace,[a] Marcos V. C. Gomes,[a] Antônio C. Katata,[b] Antônio C. Zanin,[c] Lincoln F. L. Moro,[c] Carlos R. Porfírio,[d]

[a]*PETROBRAS/CENPES/PDEDS/Gás Natural/Célula de Otimização, Av. Horácio Macedo, 950 - Cidade Universitária, Rio de Janeiro, 21949-915, Brazil.*
[b]*PETROBRAS / REVAP / Otimização, Refinaria Henrique Lage, Rodovia Presidente Dutra km 143, São José dos Campos, 12223-680, Brazil*
[c]*PETROBRAS / Abastecimento - Refino / Tecnologia de Refino / Otimização, Av. República do Chile, 65 - Sala 2102 - Centro, Rio de Janeiro, 20031-912, Brazil*
[d]*PETROBRAS / RPBC / Otimização, Refinaria Presidente Bernardes Cubatão, Praça Marechal Stênio Caio A, Lima, 1, Raiz da Serra, Cubatão, 11555-900, Brazil*

Abstract

PETROBRAS has defined Real Time Optimization (RTO) as a "High Sustainability" technology for downstream operations, due to its high economic return. Since 2001, RTO tools are being tested within the Company, either using in-house process simulators or, sometimes, using available commercial ones. This paper presents an overview of the PETROBRAS experiences on RTO, showing applications on Distillation and Fluidized Catalytic Cracking (FCC) units. Alternatives based on Sequential-modular simulators, along with reduced models (Kriging models and neural nets), as well as Equation-oriented based simulators / optimizers have been explored. The project scopes vary from covering only the Reactor / Regenerator section of a FCC unit up to a whole Crude distillation unit, including the preheat train, all distillation towers and the heat and material integration. Some of these RTO applications have been running close loop for almost 6 months, with proved expressive economical benefits. Based on the knowledge acquired during all these years, some of the future development needs for the improvement of RTO technology will be presented and discussed, as a guide for future research projects.

Keywords: real time optimization, industrial application, sequential modular simulator, equation oriented simulator.

1. Introduction

RTO technology is a powerful tool for the continuous search of the most profitable way to run petroleum and petrochemical process units. Some of the major Oil Companies, as ExxonMobil, Shell and BP to name a few, have specific groups to study it and use it to improve their business [1] [2]. Cutler and Perry [3] state that despite being a hard and complex task, its potential benefits are relevant and might provide profit increases around 6 to 10% when allied to Advanced Process Control (APC).

The task of an RTO application is to make the best of an existing process unit, adjusting its process variables for every new change of external conditions, like operational variables, feed compositions and process constraints. The RTO benefits are usually

associated with the maximization of products and minimization of the specific energy consumption and other resources, depending on the following factors [4] :

- Market availability;
- Products prices and feed costs
- Safety and environmental constraints
- Product specifications

This paper focuses on the use of RTO technologies in the optimization of petroleum downstream process units. The next section provides an overview of the structure and mechanism of RTO systems. Section 3 describes PETROBRAS experience on the development and use of in-house and commercial RTO solutions. On section 4 this paper is concluded with the discussion of the key aspects to the achievement of successful implementations.

2. RTO structure, mechanism and characteristics

Though many implementation structures are possible, RTO has been usually used as a connection between the tasks of production planning & scheduling and process control [4] [7] [8] for petroleum and petrochemical applications. The central figure of an optimization application is the mathematical model. It is expected to represent the process behavior on a wide range of operating conditions with good accuracy. It is required not only to guarantee that the predicted potential profitability matches that of the real process, but also because when the optimal solution is implemented the process constraints must not be violated. Most RTO systems used nowadays are based on rigorous, steady-state, first principles mathematical models.

The good performance of an RTO system depends on a reliable mathematical model and on reliable input data. In order to obtain that, many procedures must be executed before the economic optimization problem can be solved [4] [6] [8]:

a) Gross Error Detection
b) Steady-state Detection
c) Data Reconciliation
d) Parameter estimation.

Once that a reconciled data set and a fitted model have been obtained, the process optimization can be performed. The optimization problem usually consists of maximizing the operational profit (or minimizing costs) subject to a set of constraints. On most situations the optimization problem is posed as a non-linear programming (NLP). Most commercial applications are based on variations of the SQP (Successive Quadratic Programming) algorithm. This algorithm is also used to solve the previous Data Reconciliation and Parameter Estimation problems.

3. Real Time Optimization within PETROBRAS

Since 2004, RTO has been classified by PETROBRAS and its Strategic Downstream Committee as a "High Sustainable" technology. It means that RTO is seen as a key technology to improve PETROBRAS performance and profit, and therefore significant effort and resources will be spent on this subject. PETROBRAS implementations on RTO covered a wide range of alternatives, focusing both on profitability and on the search of the best way to deliver the technology:

- Fluid Catalytic Cracking (FCC) and Crude Distillation Units (CDU)
- proprietary and commercial process models and RTO systems
- Sequential Modular (SM) and Equation Oriented (EO) approaches [6]

3.1. First experiences

The first RTO initiatives were taken using PETROBRAS' in-house process simulator for FCC, with a small scope covering only the reactor/regenerator section. The

proprietary process model used is based on a Sequential Modular (SM) approach, many difficulties were found (see next section). Nevertheless, we were able to test the technology as well as our engineers to take a step further.

3.2. Distillation Unit / SM approach (2004)

This implementation took place at the Crude Distillation Unit (CDU) and the two Solvents Units of RECAP refinery [6].
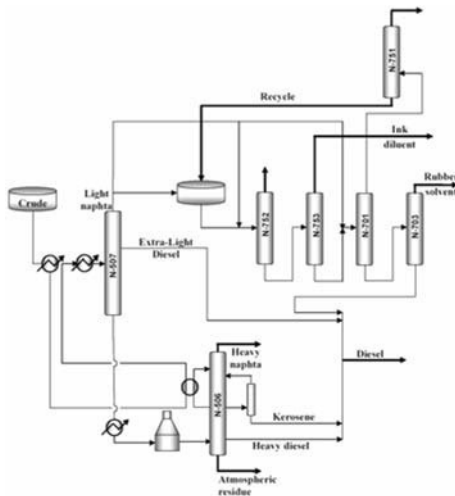
Figure 1 - **Scheme of the CDU and the Solvents Units of RECAP/PETROBRAS.**

The process model was built using PETROX, a proprietary sequential-modular process simulator from PETROBRAS. The simulation comprises 53 components and pseudo-components and 64 unit operation modules, including the 7 distillation columns and a recycle stream. All modules are built with rigorous, first-principles models.

For optimization applications, PETROX was linked to NPSOL, an SQP optimization algorithm. Procedures for Steady-state and Gross error detection, Data Reconciliation, Parameter Estimation and Economic Optimization were implemented. The economic optimization problem consisted of the maximization of the operational profit, constrained by limits related to product specifications, safety constraints, feed rate and performance parameters. The whole optimization problem involves 19 decision variables and 21 constraints.

Most of the reported problems of optimization based on sequential-modular models were observed in this application:

- Low computational efficiency, due to slow recycle loops and the numerical derivatives, that imply running the SM model several times. These derivatives are also inaccurate, which slows down the optimization process even more.
- Lack of reliability. The SM model is computed many times, and must converge always. If a single failure happens during the optimization, all the effort is lost.

In order to minimize these problems, a lot of effort must be spent on the conception, customization and tuning of the SM model. However, that is no guarantee of success. When the Data Reconciliation and Parameter Estimation problems were implemented, the same problems were observed.

3.2.1. Metamodel approach

In order to overcome some of these shortcomings, a metamodel approach has been studied. Metamodels or surrogate models [5] [6] are reduced models whose parameters are obtained with data that is generated with rigorous, first principles models. In this work, an optimization procedure was developed, combining metamodels and rigorous models with a Sequential approximate optimization (SAO) algorithm. The optimization problem is solved based on the metamodel, that is updated with data obtained from the rigorous model throughout the optimization procedure. The RECAP optimization problem was addressed with this approach, with kriging models and neural nets used as metamodels. Accurate results have been obtained with considerable reduction of the computational effort on most of the studied cases.
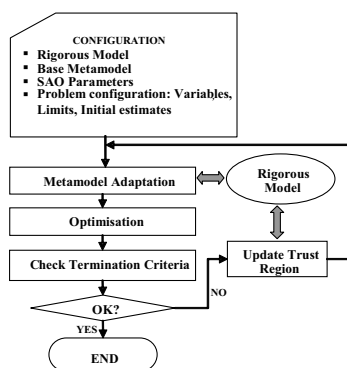
Figure 2 - SAO strategy applied to the metamodel-based optimisation.

3.3. Distillation Unit / EO (2005 to 2006)

This was the first EO RTO project PETROBRAS implemented. After an International Bid, where 3 well-known companies were invited to submit their proposals, we've decided to contract AspenTech and its AspenPlus Optimizer. The project scope included all 3 preheat trains as well as Pre-flash, Naphtha Stabilizer, Atmospheric, Vacuum and Pre-vacuum distillations towers. The unit was fully energy and mass integrated modeled on the RTO software, which allowed us to study the implications that changes on the preheat train, like feed distribution, has on the Atmospheric tower, for instance. Or to study the best pumparound heat removal distribution along this tower and its effects on the preheat train. In order to do that, all pumparounds were modeled as external streams from the tower and not as an internal model within its model (see Figure 3), as it is common on SM simulators.

The system is running on open loop since 2007. PETROBRAS intends to close loop this year after making model tuning adjustments in order to incorporate the new atmospheric trays and other unit improvements. Nevertheless, by keeping the system running open loop (around 9 runs / day), we were able to improve our knowledge of the system itself, how to overcome non convergence problems and gain expertise on how to maintain such a real time, strong data and instrumentation dependent system as well as evaluate potential benefits (around 13 000,00 dollars / day).
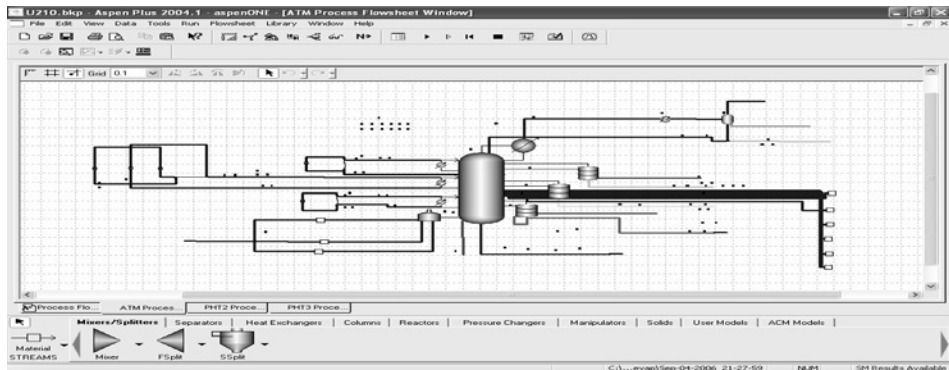
Figure 3 - Aspen Plus Optimizer Screenshot - Atmospheric tower

3.4. FCC Unit / EO (2007 to 2008)

Following the success on the distillation unit, PETROBRAS moved forward to implement an RTO on another very important unit. Again, after an international bid where the same 3 well known companies were invited to be on, we've decided to use Invensys and its ROMeo, as part of our intend to test all the available technology. The project scope included the Reactor / Regenerator section, Main Fractionator and Gas Recovery Plant. Again the unit was fully energy and mass integrated modeled.
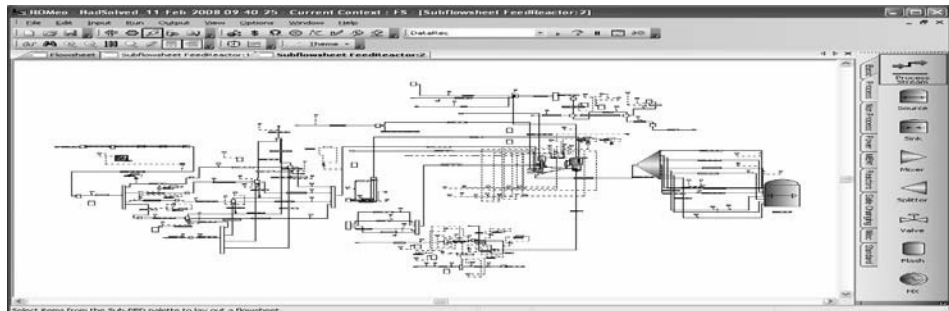
Figure 4 - ROMeo screenshot - Reactor/Regenerator Section.

The system is running on closed loop (around 8 runs / day) since June/08 with most of the independent variables active. On average, around 60% of the successful runs are being accepted by Operations and targets are being sent to Advanced Control. PETROBRAS has evaluated an average gain of US\$ 0.12 / bbl of FCC feed for this application, by comparing the unit performance with and without RTO.

A few comments on both projects:

- **Lack of instrumentation on preheat train (FCC)** – implied on simplifications, which has impacts on Main Fractionator heat balance and, thus, must be evaluated from time to time

- **Low feed lab analysis frequency** – There is a need for a better way to estimate feed characterization

- **Non-convergence problems** - Mainly, due to instrumentation faulty and/or out of service heat exchanger or other piece of equipment. Although there is a kind

of standard procedure to deal with them, it is not possible to automate it. So each problem must be solved on a case to case, hands-on basis.

These facts enforce the need for a fully dedicated RTO engineer for each application, not only to assess its results and make sure they are being implemented, but to keep the system running despite of the many daily issues the application faces.

4. Conclusions

4.1. Successful implementations

A key aspect to a successful RTO application, is the good integration among different corporate teams, like Process analysis, Production and Planning Program, Optimization and Automation, IT, Instrumentation, Laboratory, Production. A long-term plan must also be done to assure that the acquired benefits are kept and expanded, with APC and RTO experts dedicated to the on-line applications after their implementations. The process model must be updated throughout the process unit life cycle. The problem constraints and the optimization results must be continuously evaluated. The advanced and regulatory control systems are the ways by which the RTO optimal solutions are implemented. Therefore, they are essential to a successful RTO as well.

4.2. Modeling approach

PETROBRAS experiences showed that the Open-equation approach is more suitable for RTO, when compared to the Sequential-modular process models, specially when process unities of higher complexity are addressed.

4.3. Future Developments

PETROBRAS automation staff is now working to apply RTO technologies to other process technologies. Another ongoing initiative is the simultaneous application of RTO to more than one process unit.

References

[1] Terry, J.D.; Righi, J. M.; Moore, J.W.; Shah, M.K. Baytown Refinery Pipestill 3 RTO. IPS North American Client Conference, September, 2008.

[2] Greuel, J. Refinery-Wide ARPM at Sarnia Refinery. IPS North American Client Conference, September, 2008.

[3] Cutler, C. R.; Perry, R. T. Real-time optimization with multivariable control is required to maximizes profits. Computers and Chemical Engineering, 7, n.5, p.663-667, 1983.

[4] White, D. C. Online optimization: what, where and estimating ROI. Hydrocarbon Processing, p. 43-51, June, 1997.

[5] Gomes, M.V.C.; Bogle, I.D.L.; Biscaia Jr., E.C.; Odloak, D. Using kriging models for real-time process optimisation. Proceedings of the 18th European Symposium on Computer Aided Process Engineering, p. 361-366, 2008

[6] Gomes, M.V.C. Otimização Seqüencial por Aproximações – Uma aplicação em tempo real para o refino do petróleo, D.Sc. Thesis, (*in portuguese*), PEQ/COPPE/UFRJ, Rio de Janeiro, Brazil, 2007.

[7] Shobrys, D.E., White, D.C. Planning, scheduling and control systems: why cannot they work together. Computers and Chemical Engineering, v. 26, p. 149-160, 2002.

[8] Romagnoli, J.A., Sanchez, M.C., 2000, Data Processing and Reconciliation for Chemical Process Operations., Academic Press.

Embedded Control for Optimizing Flexible Dynamic Process Performance

Jeonghwa Moon, Seonbyeong Kim, Gerardo Ruiz and Andreas A. Linninger

Laboratory for Product and Process Design, Department of Chemical Engineering, University of Illinois at Chicago. Chicago, IL 60607, USA.

Abstract

High performance processes require design that operates close to design boundaries and specifications while still guaranteeing robust performance without design constraint violations. In order to safely approach tighter boundaries of process performance, much attention has been devoted to integrating design and control in which dynamic controllability as well the design decisions are taken simultaneously. However rigorous methods solving design and control simultaneously lead to challenging mathematical formulations which easily become intractable numerically and computationally. This paper introduces a new mathematical formulation to reduce this combinatorial complexity of integrating design and control. We will show substantial reduction in the problem size can be achieved by embedded control decisions within specific designs. This embedded control decisions avoid combinatorial explosion of control configuration by using a full state space model that does not require pairing of control variables and loops. The current capabilities of the methodology will be demonstrated using a realistic reactor-column flowsheet.

Keywords: Design and control integration, process and operational uncertainty, design and analysis of dynamic flexibility.

1. Introduction

The aim of integration of process design and control is finding best design decision considering economics and dynamic controllability simultaneously. During the last 30 years, several methodologies have been developed for the integration of design and control (Vassilis et al., 2004; Ramirez & Gani, 2007; Seferlis P., & Georgiadis M.C., 2004). In spite of their progress in integration of design and control field, the available methodologies still offer insufficient insight to the problems. Recently, we proposed a new method, *embedded control optimization approach* (Malcolm et al, 2007). This integrated design method can operate satisfactorily under adverse input conditions, while delivering products within desired quality specifications. In this paper, this embedded control optimization methodology is reviewed and enhanced to incorporate full state space estimator for better the dynamic process performance. A case study of designing reactor-column flowsheet is considered as a more realistic example for this methodology. In this case study, we will illustrate the ability of performing design and control integration with rigorous mathematical techniques.

2. Methodology

2.1. Problem decomposition

The conceptual problem of the integration of process design and control is a stochastic infinite dimensional mixed integer dynamic optimization problem which is extremely challenging for existing mathematical programming techniques. To overcome intractability of original problem, Pistikopoulos and co-workers proposed problem decomposition algorithms (Mohideen et al., 1996). In their problem decomposition, the main design optimization problem is solved in discretized sampling space; rigorous flexibility tests in a second stage ensure the dynamic feasibility over the entire uncertain space. If current candidate designs are not feasible in critical scenarios, these critical scenarios incorporate into the sampling space and main problem is solved again with additional augmented scenario. The final optimal designs can be achieved by repeating this iterative process. This framework is also used in our algorithm.

2.2. Embedded control optimization

Even though, the problem decomposition substantially reduces the problem size, it still remains a challenge due to combinatorial complexity of the non polynomial hard search space. Specifically, introduction of control decisions such as the insertion of feedback loops, or pairing of manipulated and control variables causes combinatorial explosion in the possible integrated design and control realizations. We therefore propose to separate the design decisions from the control decisions as shown Figure 1. At the master level, we fix design decisions such as reactor sizes, and residence time that govern dynamic process performance. No control decisions are made at this level. Once the main design decisions are specified, we assess the dynamic performance of process by using a simplified, yet reasonably competitive control schemes based on full state space identification and least square regulation. Optimal control action is calculated with relative ease based on linear state space models which are obtained dynamically in each time step. In order to keep the model consistent with the nonlinear process dynamics, the process identification is repeated in each time step. The detailed description is found elsewhere (Malcolm et al., 2007).
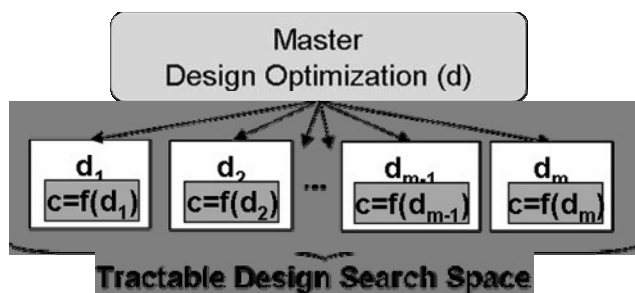
Figure 1 Proposed embedded control optimization structure. It optimizes control choice with given design.

3. Flexible design of reactor column

This section demonstrates the effectiveness of embedded control optimization by designing a reactor-column flowsheet process with uncertain reaction coefficients. This is similar to processes introduced and studied by Luyben (2007). The aim of this case study is to determine optimal design specifications with reasonable control for

dynamically flexible operations. This task of design and control integration should be done simultaneously with reasonable computational effort.

3.1.1. Process description.

This flowsheet has one reaction and one distillative separation column, as shown in **Figure 2**. The precursor A enters the reactor to be converted into the product B with an exothermic irreversible first order reaction.

$$A \xrightarrow{k_0} B \tag{1}$$

The reaction cannot be driven to full conversion, because this operation would require too high temperature causing the product B to be destroyed. In addition, this high temperature might violate safety constraints and lead to reactor explosion. To control temperature in reactor, it is equipped with a cooling jacket. Its effluent F is directed to the continuous distillation column in order to separate the product B from unreacted raw material A.
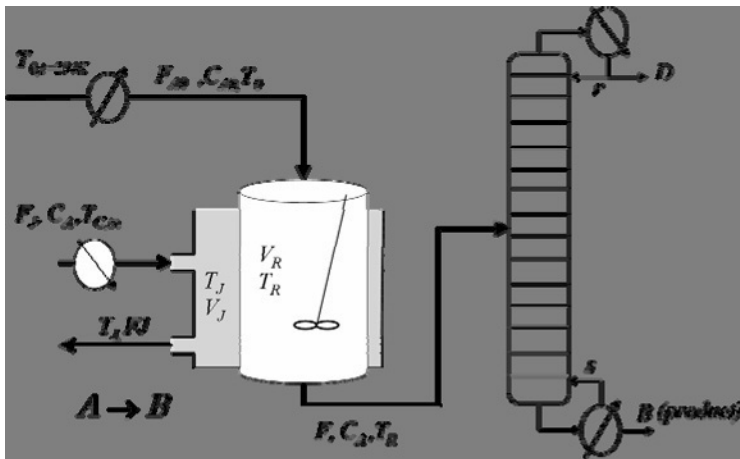
Figure 2 Reactor-column flowsheet. This is composed of a reactor and a column.
The reactor is surrounded by a cool jacket.

3.1.2. Design & manipulated variables.

Reactor diameter (D_R) and length (L_R), and heat transfer area (A_J) are treated as design variables. Also the number of stages (Ns), feed stage, and column diameter (N_f) are design variables of the column,. For manipulated variables, we select input feed temperature (T_0), coolant temperature (T_J), reflux ratio (r), and reboil ratio (s). The values of the design variables and manipulated variables need to be determined to insure safe operation within desired quality standards, under any operating conditions and in the presence of uncertain reaction conditions.

3.1.3. Operational constraints.

For safe operation, this process needs to satisfy three constraints at all times.
For the safety, reactor temperature (T_R) should never exceed 385K,

$$T_R < 385 \tag{2}$$

For the product quality, we enforce the mole fraction of component A (x_A) of final product B should be less than 0.05.

$$x_A < 0.05 \tag{3}$$

For minimum process productivity, the conversion ratio (χ) should be greater than 0.7.

$$\chi > 0.7, \quad \chi = 1 - \frac{C_A}{C_{A0}} \tag{4}$$

Thus, control variables are T_R, x_A, and χ. Among these constraints, the safety constraint needs to be enforced dynamically for all time periods of possible scenarios, as opposed to merely the steady state which is the sole concern in classical flexibility analysis. A certain design specification does not satisfy safety constraint in dynamic state, even though it does not have any violations in steady state. Therefore, we need to consider the dynamic flexibility for the optimal dynamic process performance. Productivity and quality constraints are soft constraints, which lead to performance losses that cost money, but need not to be enforced rigorously for all time intervals.

3.1.4. Uncertainty scenarios.

We wish to investigate the impact of two main uncertain parameters, associated with chemical reactions. The first parameter is preexponental factor k_0, and the second is heat of reaction λ. Their nominal values and variance are illustrated in **Table 1**.

Table 1 Nominal values and expected deviation of uncertain parameters

	θ^N	$\Delta\theta^+$	$\Delta\theta^-$	var
k_0	20.75e6	2.07e6	2.07e6	10%
λ	-17.43e3	-1.74e3	-1.74e3	10%

As a first attempt to perform the stochastic optimization, we chose 10 samples in the uncertain space of reaction conditions using Latin hypercube method, and evaluated the probabilities of each parameter set to calculate expected cost.

We now wish to rigorously determine the design and manipulated variables in the manner that, with every realization of the reaction conditions and all variability in the dynamic performance by uncertainty, the process does not violate the constraints and produces product B in desired purity limit.

3.2. Identification test

In order to employ the embedded control optimization, our methodology dynamically performs of adaptive state space identification. This repeated identification is to convert highly nonlinear process model into a simple linear state representation in open loop. With this simplified linear model, the linear regulator can find optimal control moves. The performance of process in response to step change in feed rate as well as the impact of sinusoidal disturbance is show in **Figure 3**. We have now identification phase which is occurring until time step of 40,000. In each step, there is a little change because the process is operated with approximately constant conditions, only in sinusoidal disturbances slightly impacts processes. However, a big impact occurs at t=40,000, in which the flow rate is suddenly increased to double. Now dynamics of whole process is changed and we need to adjust to new state space. This is done very satisfactorily, as shown **Figure 3**. The blue lines represent nonlinear data from mathematical model; red lines represent predicted data by identifier. Even though model is linear, because it is updated in every time step, it can maintain reasonable predictability even in a transition phase.
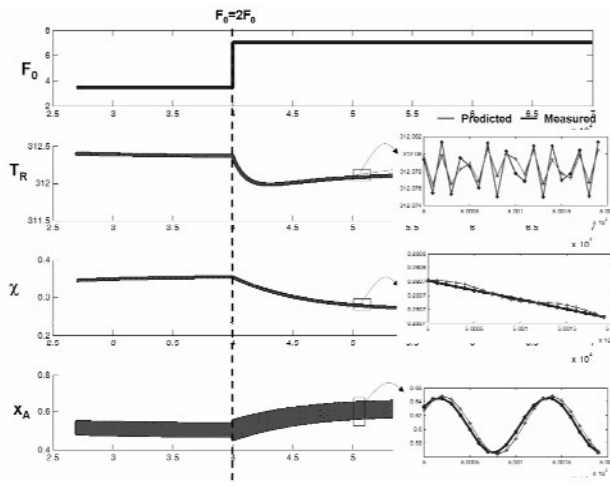
Figure *3* Performance test of identifier. This figure shows sequential least identification method predicts dynamic system behavior well, even in dynamic transition phase.

3.3. Control performance of there different designs

Next, we tested controllability of several designs under specific disturbance scenarios. Dynamic controllability of three reactors with different volumes of 12, 23, 143m^3 were investigated in response to dramatic increase of the feed rate. The doubling feed occurs at t=40,000.

Case 1: Small reactor, $V_R=12m^3$. When doubling the input feed, this small reactor is loosing the controllability- the control is not capable of handling this upset to process conditions as shown in **Figure 4**. The reactor temperature drops and reaction conversion declines, and product quality is violated.
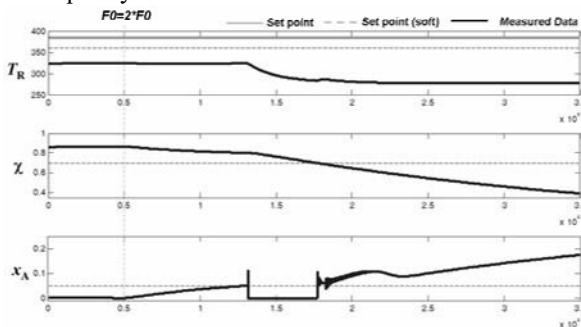
Figure *4* Control performance with VR=12m^3. It fails to keep the values of two control variables (χ, x_A) under the set points.

Case 2: Middle reactor, $V_R=23m^3$. We tested controllability of a middle size reactor. With this design specification, the controller rejected the disturbance well. The quality constraint was violated briefly in the instance of the disturbance inception, but the controller quickly removed this quality issue and kept values of all process variables under the set points. This design handles disturbance very well with simple control scheme.

Case 3: Large reactor, $V_R=143m^3$. Finally we tested larger reactor size. Our model demonstrated that the feed flow disturbances do not affect the process dynamics much in this over-dimensioned design. So the large reactor is very robust against the disturbance, but it is not competitive because of cost. It turned out that middle reactor generated more profit than the large reactor. Moreover, large reactors exhibit sluggish response and slow servo performance when set points are changed to adjust to different product specifications. This example supports the notion that arbitrary overdesign of equipment is not a solution to ensuring dynamic process flexibility of high performance processes.

3.4. Optimal integrated design with control

The previous analysis of different design demonstrated the trade off between controllability and servo performance. When considering also costs and profits what would be the best integrated design and control? For maximizing the performance, while at the same time planning flexible operation, we performed the design optimization under uncertainty as follows. For capital configuration, we consider the reactor, the column and the heat exchangers. Also we consider energy consumption for input feed temperature, cooling temperature, the reboiler and condenser, as operating cost. Also, we consider product prices to calculate total annual profit. The master level of this problem is to maximize total profit. To solve optimal design problem, Nelder-Mead simplex method is used in master level of our methodology. We found the best optimal design after 34 iterations.

4. Conclusions & future work

This paper describes conceptual framework for integrating of design and control developed by our group. Our methodology is to recast integrated design and control problem into solvable mathematical programming formulation. The case study described to show suitability of our methodology. In future, more challenging flowsheets will be examined. Also we wish to improve the quality of identification for highly nonlinear processes using more advanced identification such as subspace identification method or nonlinear model predictive control. However these advanced algorithms are more expensive, the trade off between accuracy and performance of algorithms need to be considered.

References

Luyben, W. L. (2007). Chemical reactor design and control. *Wiley-Interscience*.

Malcolm, A., Polan, J., Zhang, L., Ogunnaike, B. A., & Linninger, A. A. (2007). Integrating systems design and control using dynamic flexibility analysis. *AIChE Journal*, 53, 2048-2061.

Mohideen, M. J., Perkins, J. D., & Pistikopoulos, E. N. (1996). Optimal design of dynamic systems under uncertainty. *AIChE Journal*, 42, 2251-2272.

Ramirez, E., & Gani, R. (2007). Methodology for the design and analysis of reaction-separation systems with recycle. 2. Design and control integration. *IECR*, 46, 8084-8100.

Seferlis P, & Georgiadis M. C., The Integration of process design and control. New York: Elsevier; 2004.

Vassilis S., Perkins J. D., & Pistikopoulos, E. N. (2004). Recent advances in optimization-based simultaneous process and control design. *Comp. & Chem. Engr.,* 28(10): 2069-2086.

1257

Comparison Between Statistical and Observer-Based Approaches for Fault Detection and Isolation in a Chemical Process

Thiago de Sá Feital[a,b], Uwe Kruger[b], José Carlos Pinto[a], Enrique L. Lima[a]

[a]*Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, 68502, Rio de Janeiro, Brazil*
[b]*Department of Electrical Engineering, The Petroleum Institute, P.O. Box 2533, Abu Dhabi, U.A.E.*

Abstract

This paper summarizes the results of a comparison including three different fault detection and isolation (FDI) methods for dynamic systems. The techniques studied have recently been proposed in the literature and are Dynamic Principal Component Analysis (DPCA), Canonical Variate Analysis (CVA) and Subspace Model Identification (SMI). The aim of this study is to contrast the performance of each method in detecting and isolating incipient fault conditions. Utilizing real data from a debutaniser distillation tower, this study yields that the observer approach based on an identified SMI model is most sensitive for fault detection but performs poorly in isolating the fault condition. This method failed to correctly diagnose the fault condition using fault isolation approach. In contrast, DPCA offered a correct picture of this event using variable reconstruction and contribution charts, whilst CVA only yielded satisfactory results using variable reconstruction. For this study, it is therefore concluded that both approaches have complementary strengths and weaknesses.

Keywords: Subspace Methods, Fault Detection and Isolation, Chemical Process

1. Introduction

Based on the ever growing requirement for safe and reliably operating processes in the chemical industry, fault detection and isolation has received considerable attention over the past decades. The research literature has shown two main approaches: (i) data driven multivariate statistical techniques (Chiang *et al.*, 2001; MacGregor *et al.*, 2005) and the use of model-based state-space systems (Isermann, 2005). These approaches yielded similar fault detection and isolation methods. Given the large number of variables that are typically recorded in complex chemical, the use of mechanistic first principal models is difficult in practice, which has led to significant research activities on data-driven methods.

The aim of this article is to investigate the potential of utilizing model-based methods based on the observer design for fault detection and isolation. The required model is provided by subspace model identification, a data driven methodology that has been extensively studied over the past decade (Verhaegen, 1994; Van Overschee and De Moor, 1996). This data-driven and model-based methodology is contrasted with competitive multivariate methods for dynamic systems. Chiang *et al.* (2001) summarized that DPCA and CVA are such methods that identify a dynamic monitoring model in a reduced dimensional latent variable space.

For DPCA and CVA, process monitoring is based on the use of univariate statistics and fault isolation relies on variable reconstruction (Qin and Li, 2001) and contribution charts (Miller *et al.*, 1998). In a similar fashion to the univariate fault detection indices of DPCA and CVA, the paper proposes the use of the T^2 and SPE statistics on the basis of the estimated states and the output estimation error, respectively, for the model-based approach.

To examine the utility of the model- and observer-based approach with DPCA and CVA, the article summarizes the results of an industrial process. The objective of this study is to contrast the performance of each method in detecting fault conditions and in diagnosing the recorded event.

2. Fault Detection and Isolation Methods

This section briefly reviews DPCA, CVA and the proposed model- and observer-based approaches for FDI.

2.1. Dynamic Principal Component Analysis (DPCA)

DPCA (Chiang *et al.*, 2001) relies on a time series arrangement of a set of recorded process variables. An eigendecomposition of the covariance matrix of this augmented variable set forms then the basis for constructing a model plane and a complementary residual subspace. The retained components of this eigendecomposition relate to the most dominant eigenvalues of this decomposition and span the model plane. The residual subspace is consequently spanned by the discarded eigenvectors. Geometrically, the orthogonal projection of the augmented data vector, z, onto the model plane and the residual subspace is $\hat{z} = PP^T z$ and $\tilde{z} = (I - PP^T)z$, respectively, such that the following equality holds $z = \hat{z} + \tilde{z}$.

2.2. Canonical Variate Analysis (CVA)

Similar to DPCA, CVA performs FDI in a model plane and a residual subspace defining state sequences, x, (Chiang *et al.*, 2001). CVA relies on augmented vectors that include time lagged values of the process input and output variables. More precisely, the lagged terms are time-series arrangements of the output variables that relate to future, f, and past measurements and an arrangement of the input variables for past measurements. The state sequences are the first *d* dominant canonical variates, $\hat{x} = J_d p = U_d^T \Sigma_{pp}^{-1/2} p$, where U_d is a matrix storing the first *d* columns of U, $\Sigma_{pp}^{-1/2} \Sigma_{pf} \Sigma_{ff}^{-1/2} = U\Sigma V^T$. Moreover, p is a vector storing the past time-series arrangements of the input and output variables and the different Σ are covariance and cross-covariance matrices. The residual vector of the state-space model in term of the past is given by the *r* remaining state variables, $\tilde{x} = J_r p$.

2.3. Subspace Model Identification (SMI)

The state space model that the observer-based FDI scheme requires is provided by an SMI technique. Based on the numerically stable and efficient QR and SVD decompositions, the MOESP algorithm (Verhaegen, 1994) has gained attention. In addition, Chen (1999) reported that the subspace identification offers various practical advantages, mainly for industrial process systems, over the classical realization from input/output identified models (Chen, 1999). This algorithm relies on future and past arrangements of the input and output variables to form a total of four block Hankel matrices, where the samples are stored row wise. These matrices are then stacked prior

to the application of a QR decomposition. Together with a subsequent SVD decomposition, the quadruple of state space matrices, A, B, C and D can be extracted from the R-matrix of the QR decomposition and the left singular vectors of the SVD decomposition.

2.4. Observer Designs

The observer-based approach for FDI traditionally uses process variables to estimate output variables through a state-space model in a closed-loop design. The provided output estimation error can then be monitored as residuals (Patton and Chen, 1997). As the design of such observers may offer degrees of freedom, many approaches have been proposed to tackle uncertainties and disturbances issues. This paper uses the traditional Luenberger state observer designed by pole placement.

2.5. Univariate Statistical Fault Detection Indices

For fault detection, two traditional univariate statistics are employed for multivariate statistical process monitoring, the Hotelling's T^2 statistic:

$$T^2 = t^T \Lambda^{-1} t \tag{1}$$

which follows an F-distribution and a squared prediction error:

$$SPE = r^T r \tag{2}$$

for which approximate distributions have been proposed in Box (1954). In Equation 1, t denotes an uncorrelated mean-centered variable set and Λ, its covariance matrix.

Table 1 shows the definition of these statistics for each method.

Table 1 – Fault Detection Indices

Technique	Principal Subspace	Residual Subspace
Based on DPCA	$t^T \Lambda^{-1} t$	$\tilde{z}^T \tilde{z}$
Based on CVA	$\hat{x}^T \Lambda^{-1} \hat{x}$	$\tilde{x}^T \tilde{x}$
Based on observer	$\hat{x}^T \Lambda^{-1} \hat{x}$	$e^T e$

The work in reference Nomikos and MacGregor (1995) allows the determination of confidence limits for each of these statistics with a significance of 0.01 for example.

2.6. Fault Isolation Indices

This task relates to the determination of which process variables are mostly affected by a detected fault condition. The research literature has proposed two different concepts: (i) the use of contribution chart (Kourti and MacGregor, 1996) and (ii) variable reconstruction (Dunia and Qin, 1998). These methods are readily applicable for DPCA (Qin and Li, 2001) and CVA (Lee *et al*., 2006). The observer-based method relies on the use of generalized observers, as discussed in Isermann (2005). More precisely, each output variable is reconstructed in turn by eliminating the associated column from the observer setting and using the remaining output variables. Table 2 summarizes the isolation indices for each method in which *a* implies "affected", *i*

denotes the ith process variable and * represents a vector reconstructed by excluding the ith process variable.

Table 2 Isolation Indices

| Technique | Variable Contribution | | Variable Reconstruction |
	Principal Subspace	Residual Subspace	Residual Subspace
Based on DPCA	$t_a \lambda_a^{-1} p_{a,i} z_i$	$\tilde{z}_i^T \tilde{z}_i$	$\dfrac{\tilde{z}_i^{*T} \tilde{z}_i^*}{\tilde{z}^T \tilde{z}}$
Based on CVA	$\hat{x}_a j_{a,i} z_i$	$\tilde{x}_a j_{a,i} z_i$	$\dfrac{\tilde{x}_i^{*T} \tilde{x}_i^*}{\tilde{x}^T \tilde{x}}$
Based on observer	-	$e_i^T e_i$	$\dfrac{e_i^{*T} e_i^*}{e^T e}$

3. Industrial Application Example

This section presents a debutaniser process used to contrast the performance of each method in detecting and isolating a fault condition. The process is designed to purify a mixture of hydrocarbons into butane from a fresh feed which is product of a depropaniser distillation. Therefore, the fresh feed is not steady and can interfere in the top and bottom temperatures by an incipient change. The recorded fault data set contains a small drop in the fresh feed followed by a larger one, which some minutes later affects all the temperatures of the process. In such undesired conditions, the concentration of the head-product is altered, increasing the concentration of pentane. In order to manage the quality of this process, it is required a dynamic model and a fast FDI method.

The industrial data were recorded at a sampling interval of 30 seconds, including 8000 samples describing normal operation conditions for the training data set and 4744 samples containing the fault condition for the testing data set. The input variables are: the reflux flow (u_1), fresh feed flow (u_2) and fresh feed temperature (u_3); and the output variables are: mid-tray temperature (y_1), top-tray temperature (y_2), butane product flow (y_3), bottom-tray temperature (y_4) and inlet temperature (y_5).

4. FDI Results using DPCA, CVA and Observer-Design

This section summarizes the application of the presented fault detection and isolation methods to the recorded data sets. The subspace methods, CVA and MOESP, relied on 2 block rows, whilst for the DPCA, Parallel Analysis showed that the number of principal components does not change with the number of lag variables from 5 lag variables on. Table 3 gives the number of selected components for each method, where for CVA and MOESP it was selected according to the Scree test (Chiang et al., 2001). The observer design was based on the selection of poles inside the unit circle.

For fault detection, three parameters were used to assess the performance of each method: type I error (false alarm rate), type II error (misdetection rate) and the detection delay. Table 3 also summarizes the fault detection results and indicates that the observer-based approach outperformed the statistical-based ones in the last two parameters. Ideally, the methods should have values around 1% for the type I error due to the confidence limit chosen. Such result can be explained due to the fact that the observer-based approach has a well-defined dynamic structure, which guarantee quick and accurate responds compared to CVA-based approach. Regarding to DPCA-based

approach, which dynamic structure is based on an ARX model, the detection delay was significant once the influence of the lag variables (fault-free) in the moment of the fault can be high.

Table 3 - Fault Detection Results

Method	l	n	Type I Error (%)		Type II Error (%)		Detection Delay (min)	
			T^2	SPE	T^2	SPE	T^2	SPE
DPCA	5	7	0.7	1.3	20.7	24.8	57	51.5
CVA	1	5	0.4	2.3	27.4	22.4	84.5	50.5
OBS	1	5	0	0.1	5.4	14.4	27	34

Next, the performance of each method for fault isolation was contrasted. Table 4 summarizes the results of each method using contribution charts and Table 5, using variable reconstruction approach. Letters *a* and *b* in Tables 4 and 5 stands for the first and the second fault detection indices, respectively.

The fault isolation step via contribution charts showed divergent results among the methods. DPCA-based approach presented the best description of the fault since the residuals were able to detect the minimal increasing in the fresh feed, and the T^2 contributions could reveal the most affected variable. From this result, the root cause of the fault can be clearly identified and the fault diagnosis can be correctly performed as a trouble in the feed stream. CVA-based approach had similar result compared to DPCA only for the model subspace. However, the "faulty" variable (u_2) was missed in such way that impaired the correct fault diagnosis. Finally, the observer-based approach was unable to make a reasonable description of the fault since the proposed fault detection index for the state variables (OBSa) does not allow a kind of contribution and the residuals are restrictive only to the output variables.

Table 4 - Fault Isolation Results Via Variable Contribution

Methods	Variable Contribution							
	u_1	u_2	u_3	y_1	y_2	y_3	y_4	y_5
DPCAa	40.3	3.8	46.2	**734.2**	64.4	34.2	6.1	43.8
DPCAb	16.3	**66.6**	1.5	20.5	1.1	17.5	0.3	2.3
CVAa	0.9	1	4.9	**49.8**	0.5	2.3	21.8	8.4
CVAb	3.7	2.8	4.3	54.7	**157.1**	2.7	37.4	27.9
OBSa	-	-	-	-	-	-	-	-
OBSb	-	-	-	1576	576	263	359	**4626**

The variable reconstruction approach also had good results compared to the contribution charts, concerning the statistical methods. On the other hand, the bank of generalized observers did not show any improvement. Although this approach can make an analysis of the input variables as well as the output variables, the results of these industrial data sets showed that it does not work well. The observers designed for the input variables monitoring were able to predict correctly the output variables so that the influence from the input variables was not significant and the fault isolation turned out to be impractical.

Table 5 - Fault Isolation Results Via Variable Reconstruction

Methods	Variable Reconstruction							
	u_1	u_2	u_3	y_1	y_2	y_3	y_4	y_5
DPCAb	0.89	**0.52**	0.96	0.87	0.98	0.84	1.00	0.94
CVAb	44.80	**32.98**	45.05	36.13	42.71	44.75	34.31	42.48
OBSb	0.000	0.000	0.000	0.019	**0.005**	0.009	0.009	0.010

5. Conclusions

This paper studied the performance of the most traditional fault detection and isolation methods in a chemical process. In order to make a reasonable comparison, univariate statistics for the state variables and the output estimation errors of the observer-based approach were proposed. Results from a real process data showed a very accurate detectability in the case of the observer-based method. The dynamic structure of such method can be handled by pole placement allowing different types of respond. For the fault isolation step, only statistical-based methods were able to make a good description of the fault. Therefore, none of the methods showed a significant advantage in both fault detection and isolation steps simultaneously. A hybrid approach unifying the strengths of such methods will be the focus of future works.

References

C. Lee, S. W. Choi, I.-B. Lee, 2006, Variable reconstruction and sensor fault identification using canonical variate analysis, Journal of Process Control, 16, 747-761.

C. T. Chen, 1999, Linear Systems Theory and Design. Oxford University Press, 3rd Edition.

G. E. P. Box, 1954, Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance in One-Way Classification, The Annals of Mathematical Statistics 25, 290-302.

J. Chen, R. J. Patton, 1999, Robust Model-Based Fault Diagnosis for Dynamic Systems, Kluwer Academic Publishers.

J. F. MacGregor, H. Yu, S. G. Muñoz 1, J. Flores-Cerrillo, 2005, Data-based latent variable methods for process analysis, monitoring and control, Computers and Chemical Engineering 29, 1217–1223.

L. H. Chiang, E. L. Russell, R. D. Braatz, 2001, Fault Detection and Diagnosis in Industrial Systems, Springer.

M. Verhaegen, 1994, Identification of the deterministic part of MIMO state space models given in innovations form from input-output data, Automatica 30, 61-74.

P. Nomikos, J. F. MacGregor, 1995, Multivariate SPC Charts for Monitoring Batch Processes, Technometrics 37, 1, 41-58.

P. Van Overschee, B. De Moor, 1996, Subspace Identification for Linear Systems – Theory, Implementation and Application, Kluwer Academic Publishers

R. Dunia, S. J. Qin, 1998, A Unified Geometric Approach to Process and Sensor Fault Identification and Reconstruction: The Unidimensional Fault Case, Computers and Chemical Engineering 22, 7-8, 927-943.

R. Isermann, 2005, Model-based fault-detection and diagnosis – status and applications, Annual Reviews in Control 29, 71–85.

R. J. Patton, J. Chen, 1997, Observer-based fault detection and isolation: robustness and applications, Control Engineering Practice 5, 5, 671-682.

S. J. Qin, W. Li, 2001, Detection and Identification of Faulty Sensors in Dynamic Processes, AIChE Journal 47, 7, 1581-1593.

T. Kourti, J. F. MacGregor, 1996, Multivariate SPC Methods for Process and Product Monitoring, Journal Of Quality Technology 28, 4, 409-428.

1263

Modelling and Control of the Variable Channel Reactor

Mayank P. Patel,[a] Nilay Shah,[a] Robert Ashe,[b]

[a]*CPSE, Dept. Chemical Engineering, Imperial College, London, SW7 2AZ, UK*
[b]*AM Technology, The Heath Bus. & Tech. Park, Runcorn, Cheshire, WA7 4QX, UK*

Abstract

The full potential of a novel continuous reactor is addressed via the relationship between process design, the ability in handling the constraints and the optimisation framework. A Model Predictive Controller is developed to maximise the yield under hard input and state constraints. Simulations show that the designed control system gives high reaction yield and ensures that the temperatures inside the reactor are within a predefined limit. The challenges of an efficient start-up phase and the transition to the optimal operating point are also investigated by means of an event-driven approach.
Keywords: Process Intensification, process control, MPC, HEX reactor start-up

1. Introduction

The synthesis of fine chemicals and pharmaceuticals are generally carried out in a batch or semi-batch processes due to flexibility and versatility of the equipment over the continuous counterpart [Roberge *et al.*; 2005]. However, poor transfer of heat generated from reactions limits the application of batch equipment to products of low volumes and short life-times.

A new concept of small-scale continuous reactor, the Variable Channel Reactor (VCR) is being developed by AM Technology. The VCR is well incumbent under the umbrella of process intensification, as it simultaneously tolerates high reactant concentrations whilst minimising solvent usage. It is evident that reaching the optimal operating point when starting from a cold, empty reactor is a non-trivial matter.

Here we propose a control and event-driven start-up strategy with the objective of optimising the reactant conversion with constraints on both control and state variables.

2. The Variable Channel Reactor

The VCR is a modular construction of reactor plates. Through initial process design, inherent control of the temperature profile is achieved by altering the geometry of the channel along the process fluid flow path, as seen in Figure 1. The reaction power is thus controlled by varying the volume to heat transfer area. In this paper, a simple second order exothermic reaction is considered where C is the product of interest.

$$A + B \longrightarrow C + \text{heat}$$
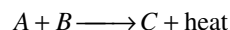
The VCR assembled as multiple consecutive channels may, from a modelling point of view, be approximated as a 1-D continuous tubular reactor with injections of reactant B along the reactor with cooling jackets around the tube. Initial experiments indicate perfect mixing conditions after a few channels, implying that the Arrhenius law can be used for modelling the reaction kinetics. From first principles, partial differential equations (PDEs) of the distributed system are discretised into elements which characterise the states of reactor temperature T and concentration C of components i.

The spatial derivatives are approximated with a first order backward finite difference method (BFDM) to a system of ordinary differential equations (ODEs);

Material balance: $\dfrac{\partial C_i}{\partial t} = -\bar{v} \cdot \dfrac{1}{L} \cdot \dfrac{\partial C_i}{\partial z} + D \cdot \dfrac{1}{L^2} \cdot \dfrac{\partial^2 C_i}{\partial z^2} + \displaystyle\sum_{j=1}^{NoReac}\left(v_{i,j} \cdot r_j\right)$ $i = 1,\ldots,$ NoComp

Energy balance: $\bar{\rho} \cdot \bar{C}_P \cdot \dfrac{\partial T}{\partial t} = -\bar{\rho} \cdot \bar{C}_P \cdot \bar{v} \cdot \dfrac{1}{L} \cdot \dfrac{\partial T}{\partial z} + k \cdot \dfrac{1}{L^2} \cdot \dfrac{\partial^2 T}{\partial z^2} + Q - Q_{rxn}$

where, $Q = \dfrac{S}{H} \cdot U \cdot \left(T_{wall} - T\right)$; $Q_{rxn} = \displaystyle\sum_{j=1}^{NoReac}\left(\Delta H \cdot r\right)$

Reaction rate: $r_j = A_j \cdot \exp^{\left(-E_a/R \cdot T\right)} \cdot \displaystyle\prod_{i=1}^{NoComp} C_i^{a_{i,j}}$ $j = 1,\ldots,$ NoReac

The terms Q and Q_{rxn} refer to the heat removal rate of the reactor element and the heat generation rate from the reaction respectively. The assumption of constant wall temperatures, T_{wall}, considers the VCR as 'defacto' isothermal. It is worth noting that the derived model is highly nonlinear due to temperature influence on the reaction rate.
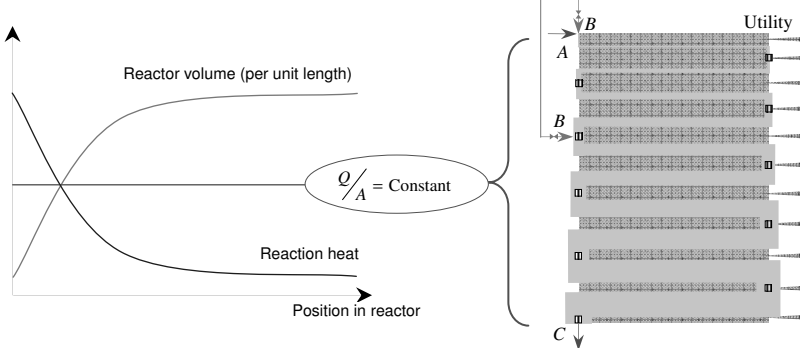
Figure 1. *Left:* The VCR concept. *Right:* A schematic of the VCR, with reactant and utility flows, i.e. 0.25 g/s and 290 g/s respectively. Temperature probes are within the vertical channels.

3. Process Operation

The main focus is on the number and locations of the injections and the cooling zones, as these significantly affect reactor performance and control conditions [Fogler; 1992]. Steady state temperature and conversion profiles are plotted for an exothermic reaction in Figure 2. Reactant *B* is distributed with 60% at the inlet and 40% in channel 9 (positioned at 40% along the reactor's length). The cooling capacity also changes with the latter injection. A maximum temperature, $T_{max}=30$ °C is in place to avoid product degradation above this limit and which inadvertently limits production capacity.
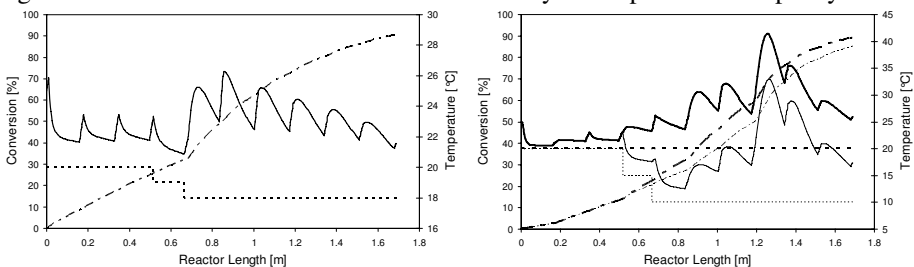
Figure 2. *Left:* Simulated process temperature profile (solid), conversion (dash-dot) and cooling (dashed) with injection at two sites giving a temperature maxima of *T*=26.3 °C. *Right:* Same as *left*, but here five injections are used with one (thick) and two (thin) cooling zones.

The importance in accurately representing the uncertainty in parameters such as reaction kinetics, micro-mixing conditions, heat transfer capacity as well as nonlinearities in the process dynamics, is crucial to give confidence in the model output [Shah *et al.*; 1999]. The overall control objective is an optimisation of the reactant conversion, γ, defined as the ratio between the amount of product formed to a theoretical conversion of 100%.

$$\max_{u \in U} \gamma(x,u,d) = \frac{C_C}{C_A + C_C} \quad \text{so that} \quad Wx \leq \Omega$$

where x are the state variables; temperatures and concentrations along the reactor, u are the three control signals, U is the admissible control set, d are exogenous disturbances with the state constraints defined as W and Ω.

4. The control strategy and design

The configuration indicated in Figure 2 is applied for simplicity, i.e. two injections of reactant B and two cooling zones. In this paper, three main control signals will be investigated; the injection flow of reactant B, u_1, and the inlet temperature of both cooling utilities, u_2 and u_3. To guarantee stoichiometric balance, the total flow rate of reactant B is fixed (at 0.25 g/s) and the control signal u_1 denotes the fraction injected at the first point. The remainder, $(1-u_1)$ is injected at the second point. The total flow rate of reactant B is thus constant at all times as are all other inputs.

Two control approaches are available for the VCR; decentralised and centralised. In decentralised control, control loops are designed independently with automatic modes to give feedback such as PID controllers. In centralised control, the feedback control is based on a multivariable process model. As a result all control inputs and measurements are computed simultaneously. An accurate process model is thus required to implement centralised control, whereas decentralised control does save on modelling effort and is easier to tune. The centralised controller in most cases improves control performance, due to its ability to consider the cross-couplings between control inputs.

4.1. Model Predictive Control

The reaction used in the simulations is the preparation of *cis*-5-norbornene-endo-2,3-dicarboxylic anhydride, a fast Diels-Alder exothermic single liquid-phase reaction. The kinetic reaction parameters at ambient temperature are; $A_j=1.1 \cdot 10^5$ L/(mol·s), $E_a=35700$ kJ/kmol and $\Delta H=-96600$ kJ/kmol.

A linear MPC controller is developed based on notations from Maciejowski [2002] and Morari *et al.* [2004]. The nonlinear model is linearised around a working point (x^0, u^0) and is sampled with $h=1$ s to a discrete-time system. The VCR is discretised in 20 elements. There are six states within each element, the reactor and two utility temperatures, with the concentrations of the reactants and product, giving the linear model 120 states.

The controlled outputs z are chosen to be the outlet concentrations of the two reactants. The cost function to be minimised is employed from Maciejowski [2002] and the reference trajectories are set to $r=[0\ \ 0]^T$, corresponding to $\gamma=100\%$.

$$V(k) = \sum_{i=1}^{H_p} \left\| \hat{z}(k+i|k) - r(k+i|k) \right\|_{Q(i)}^2 + \sum_{i=0}^{H_u-1} \left\| \Delta\hat{u}(k+i|k) \right\|_{R(i)}^2$$

The temperatures inside the reactor are subjected to hard constraints due to product degradation, $T_i \leq T_{max}=30$ °C, where T_i is the temperature in the i:th element of the model. There are also constraints on the control signals, $0.1 \leq u_1 \leq 0.9$ and 5 °C $\leq u_2; u_3 \leq 30$ °C, with the changes $|\Delta u_1| \leq 0.1$ and $|\Delta u_2; \Delta u_3| \leq 1$.

The prediction and control horizons are set to $H_p = 200$ and $H_u = 10$ respectively, as a

result the prediction time is $H_p \cdot h = 200$ seconds. The weighting matrices for the controlled variables Q and the control actions R are chosen as $Q = 10^{-5} \cdot \mathbf{I}$ and $R = [1000\ 0;\ 0\ 1;\ 0\ 1]$, so to remain below the temperature constraint of 30 °C.

The VCR is subjected to an unmeasured inlet disturbance d namely the inlet concentrations. The temperatures of the reactor walls, the inlet and outlet cooling utilities are the available state measurements, y.

The MPC controller and the nonlinear model are implemented in Matlab and gPROMS. An extended Kalman filter (EKF) is designed from a tutorial given by Dochain [2003]. The filter calculates the gain matrix $P(\hat{x})$ as the solution of the dynamical Riccati matrix equation, thus minimising the variance of the estimation error. R_1 and R_2 are the variances of the process and measurement noise, respectively.

$$\frac{d\hat{x}}{dt} = -f(\hat{x}, u, d) + K(\hat{x})(y - \hat{y}) \qquad ; \qquad K(\hat{x}) = \frac{P(\hat{x})C_y^T}{R_2}$$

$$\frac{dP}{dt} = -\frac{PC_y^T \cdot C_y P}{R_2} + PA^T(\hat{x}) + A(\hat{x})P + R_1$$

4.2. Event-driven start-up control

Upon start-up, temperatures within the VCR should guarantee that a reaction starts and that the reactants are consumed, for example at the first reactant B injection point T_1. To satisfy these two conditions, the start-up sequence is divided into four modes, schematically shown in Figure 3. With an open-loop start-up control procedure, it is possible to find the best operating point, whilst obeying the temperature constraints, However, this is difficult to isolate a suitable route in taking the process from initial rest, through to start-up to optimal operation. Therefore an event-driven start-up as described in Figure 3 is implemented with feedback control. For each step of the start-up, the already developed MPC controller due to its multivariable nature and its capacity to handle state constraints is applied.
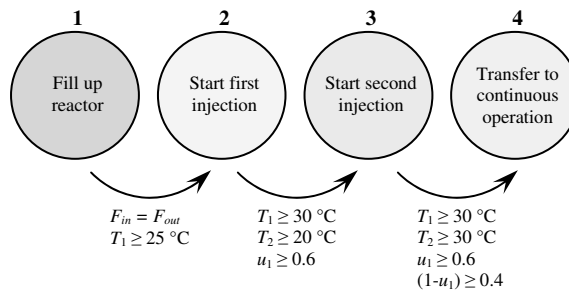
Figure 3. A finite state machine illustrates the steps during start-up.

5. Simulation results

The process starts at rest in a non-optimal working point. The rejection property of the EKF and the MPC controller is tested with a ramp increase of 5% in the concentration of both reactants during $t = 50$-100 s, see Fig. 4 and 5. This increase increases the heat released from reaction and the associated risk of violating the temperature constraint.

The signal u_1 increases from 0.6 to 0.65 within the first 50 seconds, and thus redistributing reactant B from the second injection. Temporary changes in stoichiometry cause a drop in the product yield during the first 70 s. The flow from the second injection point, i.e. 40 % along the reactor length, to the exit is around 20 s which

explains the time delay. The MPC controller optimises conversion by increasing the temperatures inside the reactor by increasing the cooling temperature u_2 and u_3 by 0.8 °C and 1.5 °C, respectively. However, since the cooling temperature has a limited impact on the reactor temperature in the first part of the reactor, the reactant injection distribution was used to increase the temperature there. After t=50 s, the temperatures throughout have reached the prerequisite constraint of 30 °C. To maximise the yield, the temperatures are kept just below this constraint level, which indeed is the main limitation to the reactor's performance.

During the disturbance, the disturbance rejection property of the EKF and the MPC controller reacts by decreasing u_1 that is, redistributing more reactant *B* to the second injection and increasing the cooling rates, see Figure 5. With the estimation, the MPC controller can ensure that the temperature constraints are only slightly violated during the ramp and afterwards in steady-state the temperatures stay on the constraint. This is the main advantage of multivariable model-based control [Qin and Badgwell; 2003].
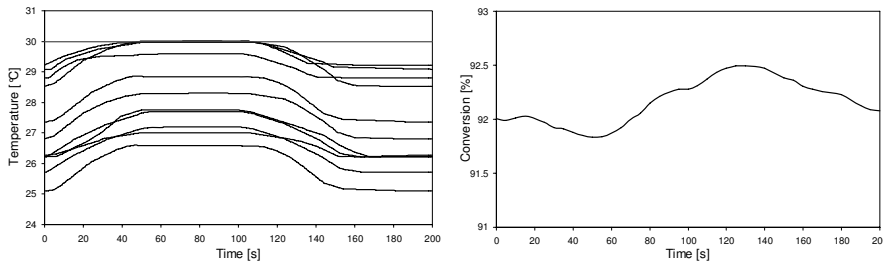
Figure 4. *Left:* Temperatures within each channel containing temperature probes whilst being under a ramp disturbance. *Right:* Outlet conversion under the 5% ramp disturbance.
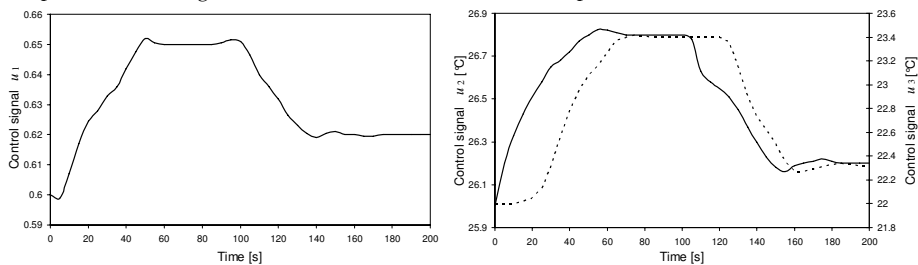
Figure 5. Control signals for the VCR. *Left:* Injection flow of reactant *B*, u_1. *Right:* Cooling utilities to the channels 1-7, u_2 (solid; right axis); and to channels 9-21, u_3 (dashed; left axis).

The closed-loop start-up control procedure is simulated with the non-linear process model is shown in Figure 6. The switch between step 1 and 2 in Figure 3 occurs at t=0, with reactant *A* is flowing through the reactor at steady-state.

Reactant *B* is injected to a channel 1 at t=0 s. As the inflow increases to its outflow the controller increases the utility to follow the temperature reference. After 3 seconds, the temperature at the first injection has reached 26 °C, meaning the reaction has started.

At t=22 s the controller switches from step 2 to 3. The inlet flow rate at the injection point in channel 9 is now included as a control variable. The temperature reference of 20 °C is tracked, and as the second feed increases, the cooling is intensified.

At t=36 s transition between step 3 and 4 occurs. At this step, the controller switches from start-up mode to continuous operation mode. The MPC now aims to optimise for product conversion, by increasing the reactor temperature to the highest allowable temperature before product degradation. In turn, some of reactant *B* is redistributed from

the second injection to the first, which decreases the need of cooling and saving energy. The transfer to optimal operation in step 4 allows the reactor to reach a conversion of 92% at the outlet, compared to 90.6% after step 3.
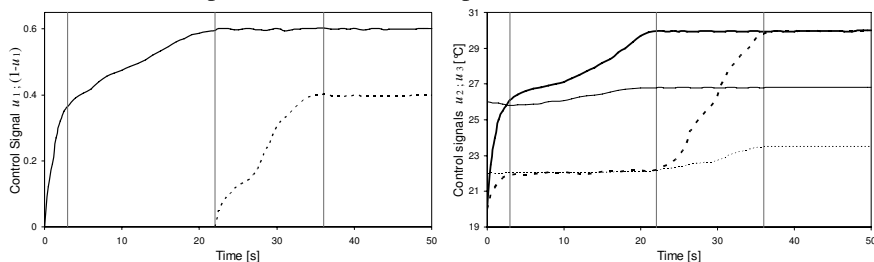
Figure 6. *Left:* During start-up, control signal of injection flow of reactant *B*, u_1 (solid) and 1- u_1 (dashed). *Right:* Temperatures of cooling utilities, u_2 (thick solid); and, u_3 (thick dashed), with temperature profiles within channels 1 and 9, T_1 (thin solid) and T_9 (thin dashed) respectively.

6. Conclusions

The concept and operation of the Variable Channel Reactor (VCR) has been presented. A combination of good mixing and high heat transfer capacity makes the VCR suitable for exothermic reactions and process intensification. The flexible configuration with cooling zones and multiple injections give possibilities for improved process control.

A centralised control is chosen over a decentralised approach, since it takes advantage of multiple inputs multiple outputs (MIMO) models, and offers explicit handling of input and state constraints. Thus, a linear MPC controller is designed, using the injection flow of reactant *B* and the inlet temperatures of cooling utilities to maximise conversion whilst obeying temperature constraints for improved product quality. An extended Kalman filter (EKF) was applied for estimation of unmeasured concentrations. Disturbance estimation was used to reduce the impact from variations in the feed concentrations. The simulations show that high reaction yield can be achieved, without violating the safety constraints. The disturbance and parameter estimation in the EKF increases the robustness of the process control system. By reducing the impact from external disturbances, the risk of unnecessary shutdowns is decreased.

The MPC controller was then applied to the start-up of the VCR which transfers the process from initial state to an operating point. Multiple steady states and multiple inlet ports along the reactor form a process with challenging dynamics. Safe and robust start-up is achieved through an event-driven sequence, i.e. divided into several steps, each associated with a transition condition that needs to be satisfied. The rule of thumb is that the feed of reactant *B* should only begin when the temperature is favourable for the reaction to start. This reduces the sensitivity to uncertainties and aids for good control.

References

D. Dochain, 2003, *Journal of Process Control*, 13, 801-818.

S. Fogler, 1992, Elements of chemical reaction engineering, Prentice Hall.

J. Maciejowski, 2002, Predictive Control with Constraints. Pearson Education Ltd.

M. Morari, J. Lee, C. Garcia, 2004, Model Predictive Control, Prentice Hall.

S. Qin and T. Badgwell, 2003, *Control Engineering Practice*, 11, 73–764.

D. Roberge, L. Ducry, N. Bieler, P. Cretton, B. Zimmermann, 2005, *Chemical Engineering & Technology*, 28(3), 318-323.

N. Shah, N. Samsatli, M. Sharif, J. Borland, L. Papageorgiou, 1999, 4th Intl. Conf. Foundations Computer Aided Process Design, Breckenbridge, USA.

1269

Development of a Strategy to Monitor and Control the Oil-Water Interface Level of a Liquid-Liquid Separator for Treatment of Wastewater Using an Image-Based Detector

Lenita S. L. Fernandes[a], João B. A. Paulo[a], Jackson A. Oliveira[a]*

[a] *DEQ - Departamento de Engenharia Química – CT, Universidade Federal do Rio Grande do Norte Campus Universitário, Lagoa Nova, 59078-970, Natal/RN – Brazil*
*Email: jackson@eq.ufrn.br

Abstract

The objective of this work is to present the development of a strategy to monitor and control the organic solvent-water interface level on the MDIF® (a mixer-settler based on phase inversion) process. This is one of the most important variables for this equipment. However, the measurement of the organic solvent-water interface level (in line) is still a hard task. There are few sensors able to measure it in laboratory scale systems due to dimensional compatibility. In the present study, a sensor for detecting the organic solvent-water interface level was developed based on the acquisition and treatment of images obtained dynamically through a camera (webcam). In addition, a control strategy was implemented in feedback mode in order to maintain the referred interface level in a desired condition. A control and data acquisition program was built in Compaq Visual Fortran language to perform the following tasks: acquisition and treatment of images for identification of water-solvent interface; decisions and implementations of control signals; and recording of data in files. Some experimental runs in open-loop were carried out using the MDIF® and random pulse disturbances were applied on the input variable (water outlet flow). The responses of interface level permitted the process identification by transfer models. From these models, the parameters for a PID controller were tuned by direct synthesis and tests in closed-loop were performed. Preliminary results for the feedback loop demonstrated that the sensor and the control strategy developed in this work were suitable for level control.

Keywords: MDIF®; process control; image-based sensor; interface level.

1. Introduction

The process imaging use with applications in process engineering has increased significantly more recently. Process information can be extracted from images to estimate important variables as part of a control scheme. Regarding the image sensors, there are three applications field with interest in image detection techniques and analysis: communication and entertainment; medical, security and scientific; and industrial sectors. In the particular case of the chemical process industry, the imaging technology can be found in a significant range of applications, such as: in three dimensional multiphase flow; in the development and control of combustion systems; in detection and measurement of particulate size and shape, mixture uniformity, amount of fluidization, process efficiency, and various factors related to product quality; in particle separation by flotation cells and columns; in thermal imaging for monitoring flow profiles, and so on (Scott and McCann, 2005). Although this technology is already in industrial use, it is still an area in development. The main objective of this work is to present a based-image sensor developed to measure, and subsequently to control, the organic solvent/water interface of a system mixer-settler based on phase inversion (MDIF®) (Fernandes Jr. et al., 2006). The idea is to promote a safer and more efficient

operation for this continuous liquid-liquid separator which is used to treat continuously wastewater of oil industries.

2. Experimental Unit - MDIF®

MDIF® (a mixer-settler based on phase inversion) is a continuous system composed basically by three sections: mixing, settling and separation (Hadjiev et al., 2004). As shown in Figure 1, in a mixing chamber (1), the water contaminated with oil (2) is put in contact with an organic solvent (3) by agitation in order to extract the oil from the water. When this mixture passes through a perforated plate (4), water drops (5) are formed and carry small oil droplets internally (6). The water, which is originally the continuous phase, becomes the dispersed phase (as drops), while the organic solvent that is added to extract the oil becomes the continuous phase of the system. The water drops descend toward to the settling chamber (7) filled with organic solvent, allowing a transfer of the oil droplets from the water drops to the continuous phase (organic solvent). In the separation section (8), the water drops coalesce and form a dense phase (treated water) containing oil in acceptable accordance with the environmental legislation in force, which, in Brazil, establishes maximum total oil and grease (TOG) of 20 mg/L. An important variable during the MDIF® operation is the organic solvent-water interface level (9) in the separation section. In order to measure this interface level, it was developed an image-based sensor, which consists on image detection by a camera (webcam) (10) and its treatment by computational routine. Using a feedback control loop, this measurement is compared to the desired level and a control action is taken according to the PID law, via computational program (11). This action is sent to manipulate the water outlet flow through a control valve (12). The interface level control is essential both to avoid the dragging of the solvent during the water removal and improve the extraction efficiency of the oil by the solvent.
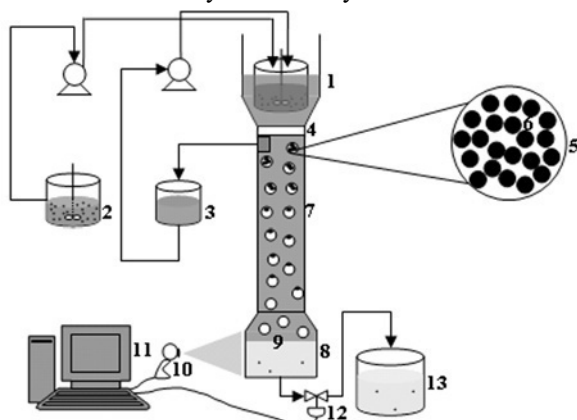
Figure 1 – MDIF® schematic representation.

3. Development of a image-based sensor

The image-based sensor was developed with the objective of detecting the organic solvent/water interface level in the MDIF® separation section. As it mentioned previously, the oil-water interface level control on the separation chamber is extremely important by two reasons: first, with regard to security and quality - the organic solvent-water interface level must be kept far from the outlet to avoid dragging organic to the treated water; second, regarding to the process performance - the organic solvent -water interface level must not ascend in the settling chamber to avoid reducing the residence time of the drops. Moreover, as the MDIF® is in laboratory scale, it is difficult to find an interface sensor with reliable dimensions to this system, and conventional level

sensors are not applied on interface detection. In this study, the image capture was carried out by a simple camera (webcam) and the interface level was estimated using a computational routine, according to the following steps:

1. The camera shows images in real time in preview mode and a static image is captured in grey scale every 3 seconds.
2. Color arrays with 200 discretization points at vertical direction from captured image are obtained in four horizontal positions and a mean is taken, as shown in the Figure 2.
3. From the color arrays mean, a theoretical curve (sigmoid function), represented by the Equation 1, is fitted dynamically through a parameters estimation method based on Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995). The first derivative of this curve is used to compute the interface level through the calculation of maximum point. This approach allows to signalize the colors transition between organic solvent phase and water phase (Figure 3).

In this method, the theoretical curve used to fit the color data of the image is given by the expression:

$$f(x) = [a - (a-b).e^{-p_1.e^{-p_2.x}}] \qquad (1)$$

where a, b, p1 and p2 are parameters; x indicates the image vertical coordinate; and f(x) is the resultant color in gray scale (0-255).
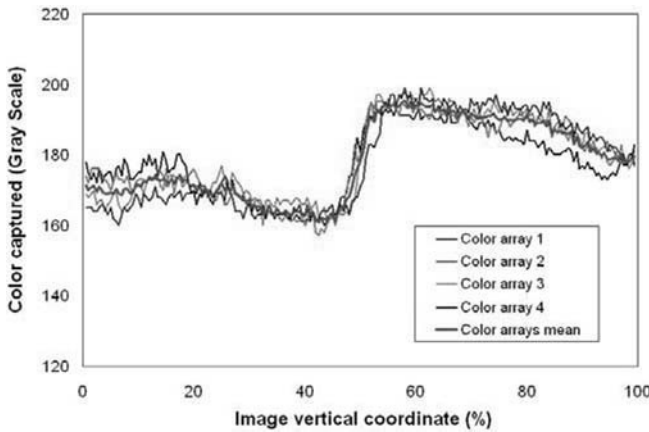
Figure 2 – Color spectra measured from the image in gray scale.
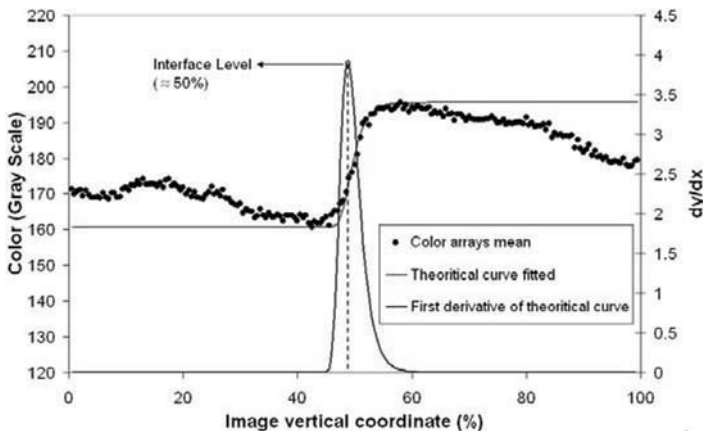
Figure 3 – Sigmoidal fitting with the color spectra mean and its first derivative.

4. Process identification and control strategy

In order to identify the dynamic model capable to describe the behavior of the MDIF® process regarding the interface level, some experiments were performed in open loop from different initial steady state. By applying random pulse disturbances with different magnitude and duration, it was possible to analyze the interface level response from the reaction curves for distinct operational conditions of flowrate. In Figure 4, it can be seen one of studied cases where an attempt of reaching the process identification was led using two transfer models: first-order and second-order functions. The parameters for each function were estimated using the Particle Swarm Optimization method (Kennedy and Eberhart, 1995). The fitted curves allow to show that the models presented a similar behavior. Therefore, it is not possible to discriminate the better model which describes the process (Figure 4). In spite of this, the transfer functions were used to find the PID (proportional-integral-derivative) controller parameters in feedback structure by direct synthesis approach (Ogunnaike and Ray, 1994), admitting a first-order closed-loop trajectory (q) with time constant (τc) of 1.0 minute. The Table 1 lists the adjusted functions for the experiment presented in Figure 4 and the parameters obtained by direct synthesis for PI and PID controllers.
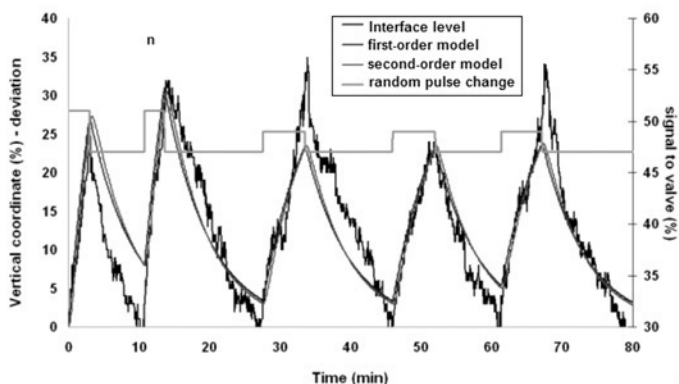
Figura 4 – Dynamic curve for the interface level as response to random pulse changes – experimental and approximate curves.

Tabela 1 – Transfer functions and tuning for feedback PID controller

First-order trajectory: $q(s) = \dfrac{1}{(s+1)}$					
System	**Approximate Model**	**Parameters**			**Controller**
		Kc	**τ_D**	**τ_I**	
First-order	$G(s) = \dfrac{18.1}{(6.32s+1)}$	0.35		6.32	PI
Second-order	$G(s) = \dfrac{17.84}{(5.78s+1)(0.34s+1)}$	0.34	0.32	6.12	PID

A control and data acquisition program was implemented with the Compaq Visual Fortran compiler (CVF 6.5) as shown in Figure 5. In this program, the following tasks are accomplished: acquisition and treatment of images for detecting the interface level; decisions and implementations of control signals based on PID controller; and recording of all data in file. In addition, as displayed in the program window (Figure 5), the users

can notice the interface level in real-time and to operate the process in manual control mode, if necessary. Other important fact is that the control strategy can be modified in order to improve the controller. In this preliminary study, only PID controller was implemented in this program. However, other control schemes can be adapted easily as routines for the program.
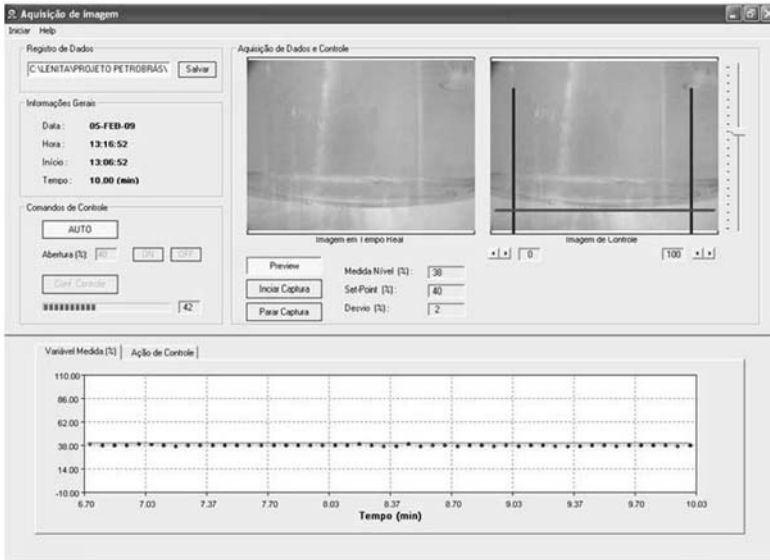
Figure 5 – Main window of the program implemented to monitor and control the MDIF® process.

5. Results and Discussion

In order to investigate the response of the closed-loop system to step changes in the set-point and in the input disturbances (Ogata, 1997), some experiments were carried out with the MDIF® process. For reason of space limit in this document only two experiments are presented here. The particular case of the set-point response for the interface level using the PID controller tuned by direct synthesis (Table 1) is shown in Figure 6. It is possible to note that the PID controller provides a response with short rise time and without offset. Although the response has presented short overshoot and damped oscillations, the controller has demonstrated satisfactory for MDIF® process.
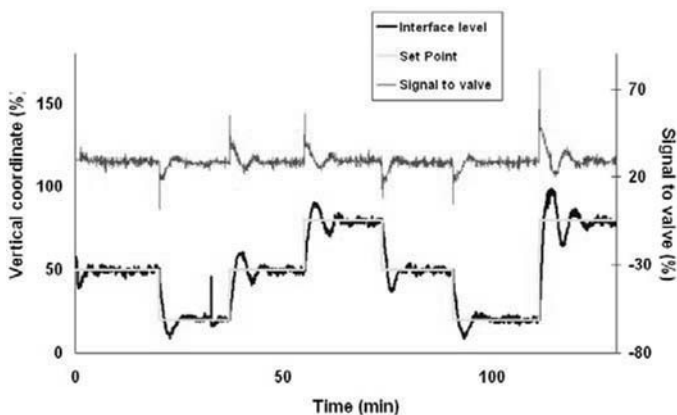
Figure 6 – Closed-loop set-point change responses for the interface level using the PID controller tuned by direct synthesis, as shown in Table 1.

In the case of response to input disturbance, the experiments were performed applying step perturbation at the process water feed through manipulating the pump flowrate. As shown in Figure 7, the step changes were done at time values signalized by dashed lines with magnitude represented by the green points. The response for this case provides overshoots when the changes are implemented. As larger the perturbation, more intense it is the overshoot. Even so, the controller is able to bring back the system to the desired set-point without offset. So, the controller presented suitable results regarding the regulatory control condition.
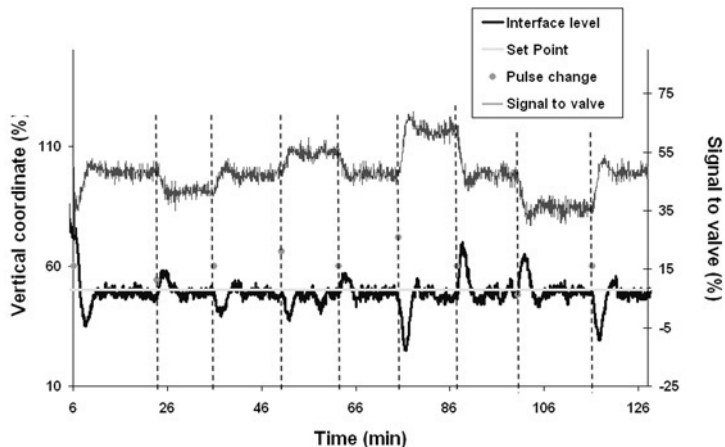
Figure 7 – Closed-loop response to the input disturbance using the PID controller tuned by direct synthesis, as shown in Table 1.

6. Conclusions

The results and analysis obtained showed that the strategy of image detection and treatment was able to measure the organic solvent/water interface level from image color spectra for the MDIF® system. The procedures to the process identification and PID controller design resulted in a closed-loop scheme that provided satisfactory responses both to servo and regulatory control. As this is a preliminary study, other controller tuning rules could be tested in order to enhance the control performance. From a practical point of view, the image-based sensor presented here may be applied as part of a control system used in other processes, where the control of interface level is required.

References

K. Ogata, 1997, Modern Control Engineering, 3ed. USA: Prentice Hall.

D. M, Scott , H. Mccann, 2005, Process imaging for automatic control. USA: Taylor & Francis group, LLC.

B. A. Ogunnaike, W. H. Ray, 1994, Process Dynamics, Modeling, and Control. Oxford University Press.

D. Hadjiev; L. Limousy; N. E. Sabiri, 2004. The design of separators based on phase inversion at low velocities in the nozzles. Separation and Purification Technology, n. 38, p. 181–189.

W. E. Fernandes Jr; J. B. A. Paulo; N. A. Moraes; A. F. Lima; G. M. Lacerda, 2006. Tratamento de águas produzidas por meio de nova tecnologia (MDIF®): aplicação para águas contendo baixas concentrações em óleo. Boletim Técnico da Petrobras, Rio de Janeiro, v.49.

J. Kennedy, R. Eberhart, 1995. Particle Swarm Optimization, in Proceedings of the IEEE Int. Conf. on Neural Networks, pp. 1942–1948.

Dynamic Degrees of Freedom for Tighter Bottleneck Control

Elvira Marie B. Aske[a,b], Sigurd Skogestad[a]

[a]Department of Chemical Eng.ineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway.
[b] StatoilHydro, Reaserch & Development, 7005 Trondheim, Norway.

Abstract

To realize maximum throughput, tight control of the bottleneck unit(s) is necessary. Dynamic degrees of freedom can be used to obtain tighter bottleneck control. Here, "dynamic" means that the degree of freedom has no steady-state effect on plant operation, like most inventories (levels). Nevertheless, temporary changes of inventories can allow for dynamic changes in the flow through the bottleneck that keeps the process closer to its bottleneck constraint and increase the throughput.

Keywords: Throughput maximization, bottleneck, inventory, ratio control

1. Introduction

In many cases, prices and market conditions are such that optimal operation is the same as maximizing plant throughput. In this case, the optimum lies at constraints, and in order to maximize throughput, the flow through the bottleneck(s) should be at its maximum at all times (Aske *et.al*, 2008). If the actual flow through the bottleneck is not at its maximum at any given time, then this gives a loss in production that can never be recovered. Tight bottleneck control is therefore important for maximizing throughput and avoiding losses.

In existing plants, the most common approach for controlling the throughput is to set the feed flow at the inlet of the plant and use inventory control in the direction of flow (Price et al., 1994). One reason for this is that most of the control structure decisions are done at the design stage (before the plant is built), where one usually assumes a fixed feed rate. However, tight bottleneck control requires that the throughput manipulator (TPM) is located close to the bottleneck (Skogestad, 2004). The term "close to the bottleneck" means that there is a short effective delay from the input (TPM) to the output (bottleneck flow).

Ideally the TPM should be located at the bottleneck, but this may not be desirable (or even possible) for other reasons. First, if the TPM is moved, the inventory loops must be reconfigured to ensure self-consistency (Aske and Skogestad, 2009). Second, there may be dynamical reasons for avoiding a so-called on-demand control structure with inventory control opposite the direction of flow (Luyben, 1999). Third, if a bottleneck(s) moves in the plant due to disturbances, then single-loop control requires relocation of TPM and reconfiguration of inventory loops. Thus, in practice one is often left with a fixed throughput manipulator, usually the feed rate. This usually leads to a large effective delay ("long loop") because the bottleneck is usually located inside the plant, and this leads to an economic loss because of a large required back off from the bottleneck constraints.

There are also related issues in business systems for using inventories as degree of freedom. Supply chains are sometimes modelled as continuous processes and Schwartz *et al.* (2006) used simulation to study decision policies for inventory management.

2. Alternative strategies for bottleneck control

Assume that the objective is to maximize the flow through the bottleneck and that the feed rate is available as a degree for freedom (TPM). Figure 1 show four ways of achieving this using simple single-loop control structures.
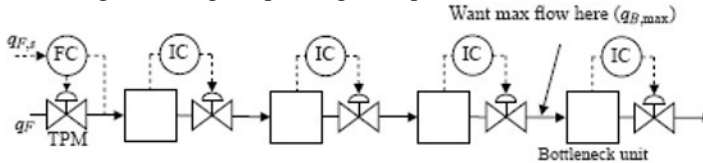
(a) **Traditional configuration** (manual control of feed rate). The feed rate is the degree of freedom for manipulating throughput (TPM), and inventory control is in the direction of flow. To maximize the flow through the bottleneck, the operators change the feed valve manually based on information about the plant operation and experience.
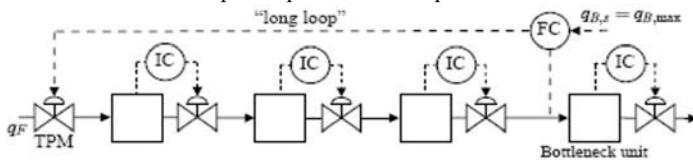
(b) **Alternative 1: Single-loop control of bottleneck flow using the feed rate.** *Problem:* The "long loop" gives a large effective delay from the feed flow (input) to the bottleneck flow (output), so tight control of the bottleneck flow is difficult.
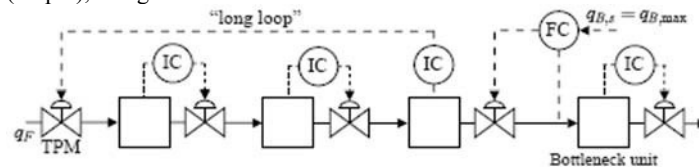
(c) **Alternative 2: Move TPM from feed to bottleneck**. This achieves tight control of the bottleneck flow. The inventory loops are not reconfigured, so the feed rate now needs to take over the "lost task" which in this case is control of the inventory upstream of the bottleneck. *Problem:* The "long loop" gives a large effective delay so control of the "lost task" may be poor.
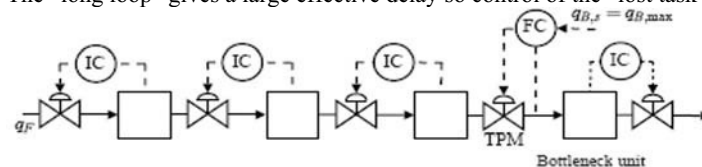
(d) **Alternative 3: Reconfigure inventory control**. The TPM is moved to the bottleneck and at the same time all the upstream inventory loops are reconfigured to be in the opposite direction of flow. Tight bottleneck control (of both flow and local inventory) may be achieved. *Problem:* Reconfiguration of inventory loops is usually very undesirable from a practical point of view.

Figure 1: Simple single-loop control structures for maximizing bottleneck flow in serial process. IC stands for inventory controller (e.g. level controller).

In summary, none of the alternatives in Figure 1 are desirable. To improve control and keep the flow through the bottleneck closer to its maximum at all times, we would like to have additional degrees of freedom, and the only ones that are normally available are the inventories (holdups) in the buffer tanks, which can be used to make

dynamic flow changes. The word "dynamic" is used because most inventories have no steady-state effect on plant operation.

The main idea is as follows: To change the flow through the bottleneck, for example, to increase it, we temporarily reduce the inventory in the upstream holdup volume. However, this inventory needs to be kept within bounds, so if we want to increase the bottleneck flow permanently, we need to increase the flow into this part of the process and so on, all the way back to the feed (throughput manipulator). The simplest (but not generally optimal) approach is to use a "ratio" control system where all flows upstream the bottleneck are increased simultaneously by the same relative amount. The idea is illustrated in Figure 2.
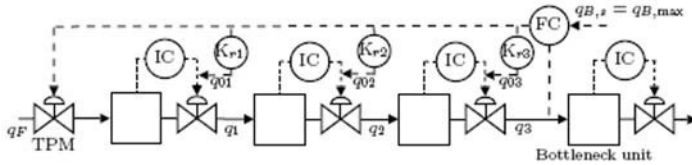
(a) **Alternative 1D: Single-loop plus ratio control.** The idea is to control the bottleneck flow by simultaneously changing all the flows upstream of the bottleneck by the same relative amount. The advantage is that the effective delay from the feed to the bottleneck may be significantly reduced and even eliminated in some cases. However, the dynamic flow changes are counteracted by the inventory controllers. In particular, note that the feed flow is the only degree of freedom that has a steady-state effect on the bottleneck flow. The strategy may also be viewed as a "ratio feedforward controller" from the feed flow to the downstream flows.
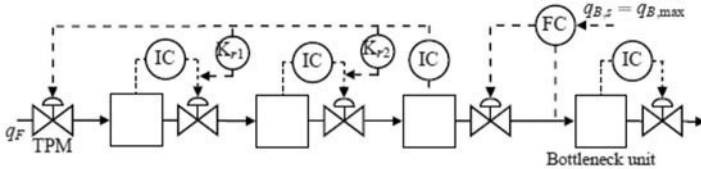
(b) **Alternative 2D: Move TPM to bottleneck and add ratio control to "lost task"**. The TPM is moved to the bottleneck and the "lost task" (inventory upstream the bottleneck) is controlled by the feed rate. The use of ratio control is the same as for Alternative 1D. The effective delay from the feed rate to the lost task is reduced by using ratio control.
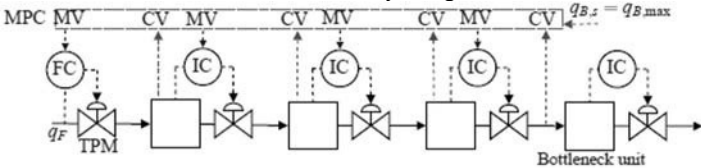
(c) **Alternative 4: Multivariable controller**. A multivariable controller (e.g. MPC) uses the feed rate and the inventory controller set points as manipulated variables (MVs). The controlled variables (CVs) are the bottleneck flow and inventory constraints.

Figure 2: Structures for controlling bottleneck flows that use inventories as dynamic degrees of freedom (with no reconfiguration of the inventory loops). Alternative 1D is studied in this paper. IC stands for inventory controller (e.g. level controller) and K_{ri} is a constant gain (ratio controller).

The most obvious is to adjust the inventory set point I_s, but it is more direct in terms of flow changes to adjust the bias q_0. The two approaches are not very different, because a change in q_0 can equivalently be implemented as a set point change by choosing $I_s = -q_0/K(s)$, where $K(s)$ is the feedback controller. In this paper, we choose to use the bias q_0 as the dynamic degree of freedom for ratio control. The important point to note is that

there are no dynamics in K_r. This means that all the flows q are changed simultaneously when q_F changes.

3. Example: Four distillation columns in series

Consider four distillation columns in series, as shown in Figure 3. The four columns represent the liquid upgrading part of a gas processing plant and consist of a deethanizer, a depropanizer, a debutanizer and a butane splitter. Assume that the butane splitter is the bottleneck unit. The throughput is manipulated at the feed to the first column. The idea is to use the column inventories (sump or condenser drum holdup) as dynamic degrees of freedom to obtain tighter bottleneck control.

The distillation column models are implemented in Matlab/Simulink. Each of the four columns is modelled as multicomponent distillation with one feed and two products, constant relative volatilities, no vapor hold-up, constant molar flows, total condenser and liquid flow dynamics represented by the Francis weir formula. All columns use the "LV-configuration" where distillate (D) and bottoms flow (B) are used for inventory control (M_D and M_B). To stabilize the column composition profile, all columns have temperature control in the bottom section by manipulating the boilup. The column inventories M_D and M_B are controlled with P-controllers with gain $K_c = 1/\tau_V$. Here we use "smooth" level control where we set $\tau_V = V_{tank}/q_{out}$ (Skogestad, 2006) where q_{out} is the flow out of the volume (D or B). The temperature controllers (TC) are tuned with SIMC PI-tuning (Skogestad, 2003) with $\tau_c = 0.5$ min.
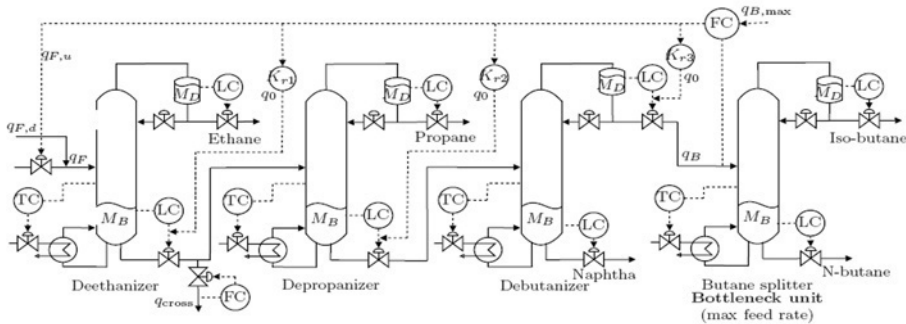
Figure 3: Distillation process: Four columns in series, here shown with throughput controlled by using single-loop with ratio control (Alternative 1D).

Two disturbances are considered. First, at *t = 10 min*, we make a 5% increase in the bottleneck flow set point ($q_{B,s}$). Second, *at t = 210 min*, there is an 8% unmeasured decrease in the feed rate to the deethanizer (q_F). The net feed flow is $q_F = q_{F,u} + q_{F,d}$, where $q_{F,u}$ is the flow contribution from the controller (initially $q_{F,d} = 0$ and $q_F = q_{F,u} = 100$, but then $q_{F,d} = -8$ at *t = 210*).

Four different control structures are tested for maximizing throughput:

1. **Manual:** Traditional (manual) control of the throughput. We assume that a skilled operator can immediately change the feed rate to the value corresponding to the new bottleneck flow set point. However, we assume that the operator does not notice the unmeasured feed flow disturbance, so no adjustment is therefore done for the feed rate disturbance.

2. **Single-loop:** Single-loop control where the bottleneck flow is controlled using the feed rate (Alternative 1).We want smooth tuning to avoid overshoot and "aggressive" use of the feed valve. Therefore, the bottleneck flow controller (FC) is tuned with SIMC tunings with $\tau_c = 3\theta$ for smooth tuning (Skogestad, 2006).

3. **Single-loop with ratio:** Use of the inventories as dynamic degrees of freedom by adding a bias (q_0) to the inventory controller outputs as in Figure 3 (Alternative 1D). In this case there is no effective delay and the bottleneck flow controller (FC) is tightly tuned with a short integral time, which are typical FC tuning parameters.
4. **Multivariable:** MPC with the feed rate and the inventory set points as MVs and the bottleneck flow and level constraints as CVs (Alternative 4). The built-in MPC toolbox in Matlab is used and tuned with a low penalty on the use of inventories (MV moves) and a high penalty on the deviation from the bottleneck flow set point (CV set point).

The four control structures are evaluated in terms of how tightly the bottleneck flow (q_B) is controlled in spite of disturbances. The resulting bottleneck flow (q_B), the net feed flow (q_F) and the inventories used as dynamic degrees of freedom (deethanizer M_B, depropanizer M_B and debutanizer M_D) for the four different control structures are displayed in Figure 4. The bottleneck control is significantly tighter with ratio control and MPC where inventories are used as dynamic degrees of freedom. The inventories are quite tightly controlled with surprisingly small variations. This follows because the disturbances introduced here are small compared to what the IC's are tuned to handle.
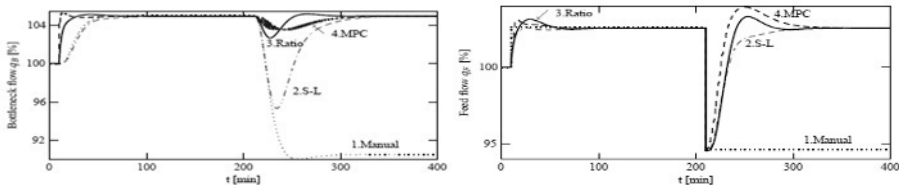
(a)   Response in bottleneck flow q_B (CV)   (b)   Responses in feed flow q_F (MV)
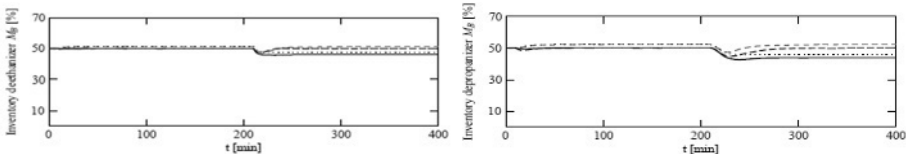
(c) Responses in deethanizer bottoms level M_B (d) Responses in depropanizer bottoms level M_B

(e) Responses in debutanizer distillate level M_D

Figure 4: Bottleneck control of the distillation process for four different control structures. 1) Manual control (dotted), 2) Single-loop control (dash-dotted), 3) Single-loop with ratio (solid), 4) MPC using both feed rate and inventories as MVs (dashed).

4. Summary: Implications for design of inventory tanks

The effect of using inventories as dynamic degrees of freedom on the design of inventory tanks is summarized. The derivations are given in details in Aske (2009).

4.1. Tank size

A desired change in tank throughput ΔqB results in a volume variation ΔV and we have

$$|\Delta V| = \tau_G \cdot |\Delta qB| \tag{1}$$

where τ_G is the time constant for "refilling" the tank. In practice, τ_G is the time for the flow rate into V to reach 63% of its steady-state change following a step in flow rate out of the (closest) upstream inventory. This is for the normal case when the TPM is upstream the bottleneck; the same formula applies also when it is downstream. For design purposes, the flow change $|\Delta q_B|$ is the (steady-state) flow change through tank resulting from the largest expected throughput (bottleneck flow) change.

Equation (1) is useful for sizing the tank (inventory volume). In words, the expected volume variation for an inventory used for bottleneck control is approximately the expected variation in flow through the unit multiplied by the time constant for the flow dynamics for "refilling" V from the upstream inventory.

4.2. Level control tuning

The level control tuning involves the closed-loop time constant (τ_V) for the level control loop in the inventory tank. We get

$$|\Delta V| = \tau_V \cdot |\Delta q_d| \tag{2}$$

where Δq_d is the flow rate change through the tank in question. Equation (2) can be used to tune the level controller, and then gives the well-known formula for smooth (averaging) level control. To see this, note that for a nominally half-full tank we must require $|\Delta V_{peak}| < 0.5\ V_{tank}$ to avoid overfilling or emptying. If we furthermore assume that the maximum expected change in flow through the tank is 50% of the nominal flow, then $q_d = 0.5\ q$. This gives $\tau_V < V_{tank}/q$, which is the well-known value for smooth level control, (e.g. Skogestad (2006)).

References

Aske, E., 2009, Design of plantwide control systems with focus on maximum throughput. Ph.D. thesis, Norwegian University of Science and Technology, Trondheim, Norway, available at www.nt.ntnu.no/users/skoge/publications/thesis/2009_ aske/.

Aske,E., Skogestad, S., 2009. Self-consistent inventory control. Ind. Eng. Chem.Res.Submitted.

Aske, E., Strand, S., Skogestad, S., 2008. Coordinator MPC for maximizing plant throughput. Comput. Chem. Eng. 32 (1-2), 195–204.

Luyben, W., 1999. Inherent dynamic problems with on-demand control structures. Ind. Eng. Chem. Res. 38 (6), 2315–2329.

Price, R. M., Lyman, P. R., Georgakis, C., 1994. Throughput manipulation in plantwide control structures. Ind. Eng. Chem. Res. 33, 1197–1207.

Schwartz, J., Wang, W., Rivera, D., 2006. Simulation-based optimization of process control policies for inventory management in supply chains. Automatica 42, 1311–1320.

Skogestad, S., 2003. Simple analytic rules for model reduction and PID controller tuning. J. Proc. Control 13, 291–309.

Skogestad, S., 2004. Control structure design for complete chemical plants. Comput. Chem. Eng. 28, 219–234.

Skogestad, S., 2006. Tuning for smooth pid control with acceptable disturbance rejection. Ind. Eng. Chem. Res. 45, 7817–7822.

Skogestad, S., 2007. The dos and don'ts of distillation column control. Trans.IChemE, Part A 85 (A1), 13–23.

A Global Optimization Approach for the Estimation of Domains of Attraction

Luis G. Matallana, Aníbal M. Blanco, J. Alberto Bandoni
PLAPIQUI (UNS - CONICET), Camino "La Carrindanga" – km 7
(8000) Bahía Blanca, ARGENTINA

Abstract

In order to completely characterize an asymptotically stable equilibrium point, some information about the size and shape of its Domain of Attraction is required. In this contribution a global optimization approach is proposed to estimate domains of attraction of general nonlinear dynamic systems. The technique is illustrated by a two states system that presents a very rich nonlinear behavior and is then applied to a typical continuous stirred tank reactor.

Keywords: Nonlinear dynamic systems, Domains of attraction, Global optimization

1. Introduction

The Domain of Attraction (DOA) of an asymptotically stable equilibrium point represents the portion of the state space where trajectories that converge to the equilibrium point originate. For the general nonlinear case, the DOA is a complicated set that does not admit analytical representation. Many techniques have been proposed so far to address the estimation of DOAs. In particular, the Lyapunov stability theory provides the basis of a family of techniques whose rationale is to approximate the DOA by a level set of a Lyapunov function. Several approaches have been proposed to identify the best level set of a given Lyapunov function by solving an optimization model. This problem is very challenging for the general case since the resulting optimization model is nonlinear semi-infinite. Formulations that make use of results on deterministic global optimization were recently proposed (Hachicho, 2007). The technique allows the identification of the best possible level set of a rational Lyapunov function that constitutes an estimation of the DOA of the equilibrium under study for nonlinear dynamic systems of polynomial type. In this contribution, an extension of the formulation presented in Hachicho (2007) is proposed. An optimization model is formulated, which includes additional constraints to avoid possible dummy solutions and makes use of a global optimization approach to address nonlinear systems of the general type.

2. Estimation of DOAs

Consider the following autonomous nonlinear dynamic system:

$$d\mathbf{x}/dt = \mathbf{f}\left(\mathbf{x}(t)\right) \tag{1}$$

being the origin of the state space, $\mathbf{x} = \mathbf{0}$, an asymptotically stable equilibrium point. The DOA of the origin is defined by the set of initial conditions $\mathbf{x}_0$, such that the trajectories initiated at $\mathbf{x}_0$ (denoted by $\mathbf{x}(\mathbf{x_0}, t)$) converge to the origin as time increases:

$$\mathrm{DOA}\left(\mathbf{0}\right) = \left\{\mathbf{x}_0 : \mathbf{x}\left(\mathbf{x_0}, t\right) \to 0 \text{ as } t \to \infty\right\} \tag{2}$$

Consider that a function *V(x)* exists such that:

$$V(\mathbf{x}) \succ 0 \text{ in } R(\mathbf{0}) \tag{3a}$$

$$dV(\mathbf{x})/dt \prec 0 \text{ in } R(\mathbf{0}) \tag{3b}$$

where R(**0**) is a region around the origin. Equation (3a) means that *V*(**x**) is positive definite in region R(**0**) and (3b) that its time derivative is negative definite in that region. If (3a) and (3b) hold for system (1) then *V(x)* is known as a Lyapunov function of (1). Let *V(x)* = c be the level set of *V(x)* at value *c*, this is, the projection on the state space of *V(x)* at value *c*. Consider the region S(**0**)={**x**: *V*(**x**) - *c* ≤ 0} is contained in R(**0**):

$$S(\mathbf{0}) \subset R(\mathbf{0}) \tag{4}$$

then S(**0**) belongs to the DOA of the origin and can be considered an estimation of DOA(**0**). A powerful technique based for the design of "maximal" Lyapunov functions has been proposed in Vannelli and Vidyasagar (1985). A maximal Lyapunov "candidate" is a rational function with the following structure:

$$V(\mathbf{x}) = \frac{N(\mathbf{x})}{D(\mathbf{x})} = \frac{\sum_{i=2}^{n} R_i(\mathbf{x})}{1 + \sum_{i=1}^{n-2} Q_i(\mathbf{x})} \tag{5}$$

Where *N(x)* and *D(x)* are polynomials defined through the $R_i(\mathbf{x})$ and $Q_i(\mathbf{x})$ homogeneous functions of degree *i* and n ≥ 2. If the system of nonlinear equations is represented in terms of a series of homogeneous functions of degree *i*, $F_i(\mathbf{x})$:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}[\mathbf{x}(t)] = \sum_{i=1}^{\infty} F_i(\mathbf{x}) \tag{6}$$

a Lyapunov function can be obtained by the procedure proposed in Vannelli and Vidyasagar (1985) and then an estimation S(**0**) of DOA(**0**) can be calculated by finding the largest positive value of *c* that verifies (4). The calculation of the maximum level set of a given Lyapunov function can be stated as:

$$\max_{c,\mathbf{x}} \quad c$$

$$s.\, t. \quad V(\mathbf{x}) = c \qquad \text{a)} \tag{7}$$

$$\dot{V}(\mathbf{x}) \prec 0 \qquad \text{b)}$$

In order to handle the negative definiteness constraint (7b), Hachicho (2007) proposed to reformulate problem (7) as:

$$c = \min_{\mathbf{x}} \quad V(\mathbf{x}) \tag{8}$$

$$s.t. \quad \dot{V}(\mathbf{x}) = 0, \quad \mathbf{x} \neq \mathbf{0}$$

The rationale behind problem (8) is to find the minimum level set of function *V*(**x**) which is contained in the level set *dV*(**x**)/*dt* = 0. The solution of algebraic model (8) is a single point in the state space, which corresponds to the tangential contact of level sets *V*(**x**) = *c* and *dV*(**x**)/*dt* = 0. In order to graphically illustrate this situation consider dynamic system (9) and Lyapunov function (10).

$$\frac{dx_1}{dt} = -x_2$$

$$\frac{dx_2}{dt} = x_1 - x_2\left(1 - x_1^2\right)$$

(9)

$$N(\mathbf{x}) = 0.593x_1^2 - 0.364x_1x_2 + 0.437x_2^2 - 0.1253x_1^4 + 0.2885x_1^3x_2 - 0.0537x_1^2x_2^2$$
$$\quad + 0.0581x_1x_2^3 - 0.0196x_2^4$$

$$D(\mathbf{x}) = 1 - 0.0001x_1 + 0.0001x_2 - 0.2685x_1^2 + 0.3217x_1x_2 - 0.1163x_2^2$$

(10)

In Fig. 1a) it is shown the Lyapunov function and its time derivative. In Fig. 1b) the level sets of interest are depicted. The closed curve $V(\mathbf{x}) = 2.2086$ (blue line) contained in the level $dV(\mathbf{x})/dt = 0$ (red line) constitutes the desired solution of problem (8) for the system under study (9)/(10). It should be noted that, due to symmetry, there are two points in the state space for the same $V(\mathbf{x}) = c$ for this system. Since both solutions are equivalent regarding the resulting estimation only one of them is reported.

3. Proposed Approach

Problem (8) is a nonlinear optimization model and therefore it may have many local solutions. A local solution may not correspond to a proper estimation of the DOA for the system under study. Consider for example the different solutions of problem (8) for system (9)/(10) shown in Fig. 2. In Fig. 2 (a to d) different level sets for the Lyapunov function are shown in blue line. In all cases the level set of the time derivative at value zero is shown in red line. The solution depicted in Fig. 2 a) is a solution of problem (8) with a value of c lower than that of the desired estimation (Fig. 1 b)). The solution corresponds to an intersection between the level set $V(\mathbf{x}) = 0.1$ and the level set $dV(\mathbf{x})/dt = 0$. It can be noticed that the intersection takes place far from the origin and therefore a very small portion of the level set of $V(\mathbf{x})$ is in fact an estimation of the actual DOA (the small circle around the origin). The solution of Fig. 2 b) verifies tangential contact between the level sets. The situation is pretty the same as in the previous case. The tangency between level sets takes place far from the origin and the estimation of the DOA is also poor. It can be observed in this case, that although the constraints of problem (8) hold, as well as a tangency condition, both $V(\mathbf{x})$ and $dV(\mathbf{x})/dt$, verify a transition of definiteness in an intermediate portion of the state space. Particularly $V(\mathbf{x})$ is positive definite in the small circle around the origin, then becomes negative definite in an intermediate region and recovers positive definiteness far from the origin. Similarly $dV(\mathbf{x})/dt$ is negative definite close the origin, becomes positive definite in an intermediate region and recovers negative definiteness far from the origin. While the small circle is an estimation of the DOA since it is fully contained in the region of negative definiteness of $dV(\mathbf{x})/dt$, the fact that the actual solution occurs beyond the transition of definiteness of the functions is clearly undesirable. Figs. 2 c) and d) are also solutions of problem (8) that present tangency between the level sets $V(\mathbf{x}) = c$ and $dV(\mathbf{x})/dt = 0$. However, none of them are actual DOAs since the level sets of $V(\mathbf{x})$ are not fully contained in the region of negative definiteness of $dV(\mathbf{x})/dt$. Since all solutions of Fig. 2 are valid solutions of problem (8) but none is the desired estimation of the actual DOA, they are considered dummy solutions of for system (9)/(10). From the above qualitative description it can be concluded that for a certain level set $V(\mathbf{x}) = c$ be an optimal estimation of the DOA, it should verify the following conditions:
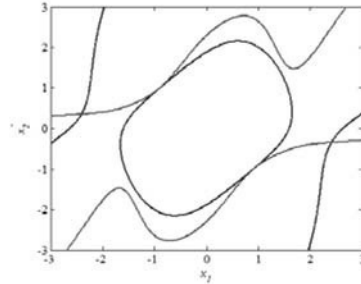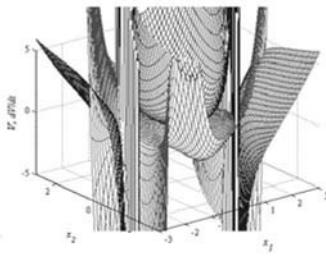
a) Surfaces: b) Level sets:
$V(\mathbf{x})$ (blue), $dV(\mathbf{x})/dt$ (red) $c = 2.2086$, $x_1 = -0.9079$, $x_2 = 0.9887$

Fig. 1: Lyapunov function and time derivative for system (9)/(10)

1. The level sets $V(\mathbf{x}) = c$ and $dV(\mathbf{x})/dt=0$ should verify tangential contact
2. The solution should belong to the portion of the state space around the origin, previous a transition of definiteness of $V(\mathbf{x})$ and $dV(\mathbf{x})/dt$ takes place
3. Provided that conditions 1 and 2 hold, the level set $V(\mathbf{x}) = c$ should be the global minimum

Condition 1 helps in avoiding dummy solutions such as the one in Fig. 2a). Condition 2 ensures that solutions of the type of Fig. 1b) are preferred over a solution like the one of Fig. 2b). Finally, the global minimum condition 3 favors the solution of Fig. 1b) over the tangential solutions of Figs. 2 c) and d). While problem (8) explicitly implies global optimality (condition 3), neither condition 1 nor condition 2, are considered. Therefore, dummy solutions such as those of Fig. 2 may result from problem (8). The proposed formulation is presented in Eq. (11) where ε is an auxiliary variable. While condition 1 can be straightforwardly imposed as a geometric constraint (11 b), condition 2 does not posses obvious mathematical expressions to be included within the optimization problem. In this contribution it is proposed to include a strict positive constraint on the denominator of function (5) to account for condition 2 (11 c). In Fig. 3 it is shown in gray the region determined by $D(\mathbf{x}) > 0$. Such condition contains the desired solution (Fig. 3 a) but excludes the dummy solution shown in Fig. 2b as can be appreciated from Fig. 3b).

$$c = \min_{\mathbf{x}} \ V(\mathbf{x})$$

$$s.t. \ \dot{V}(\mathbf{x}) = 0, \quad \mathbf{x} \neq \mathbf{0} \qquad \qquad \text{a)}$$

$$\nabla\big(V(\mathbf{x}) - c = 0\big) = \varepsilon \nabla\Big(\dot{V}(\mathbf{x}) = 0\Big) \qquad \text{b)}$$

$$D(\mathbf{x}) > 0 \qquad \qquad \text{c)}$$

$$(11)$$

4. Application Example

Formulation (11) is applied in this section to the determination of the DOA of a typical continuous stirred tank reactor. Consider the dynamic system of a typical CSTR described by mass and energy balances (12). Definitions and data are provided in Table 1. An estimation of the DOA of the steady state is calculated by solving problem (11) with Lyapunov function (13) ($x \equiv C_A$, $y \equiv T$).
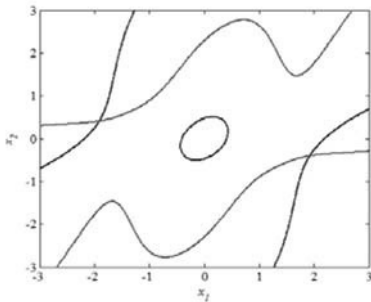
a) $c = 0.1$, $x_1 = -1.9183$, $x_2 = 0.4052$
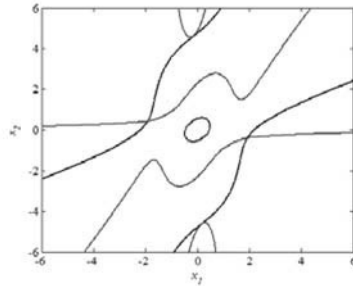
b) $c = 0.1287$, $x_1 = 0.1505$, $x_2 = -4.6180$
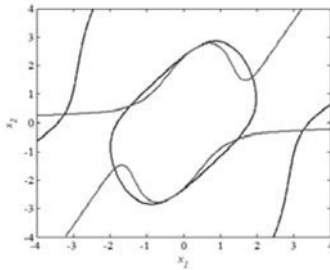
c) $c = 4.7652$, $x_1 = -0.2962$, $x_2 = -2.5906$
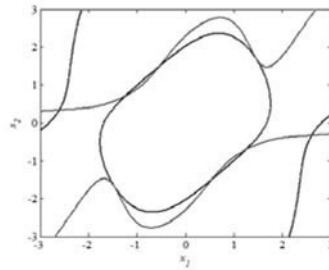
d) $c = 2.8$, $x_1 = -1.4649$, $x_2 = -1.6832$

Fig. 2: Dummy solutions for system (9)/(10)
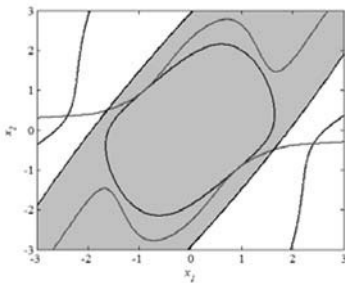
a) Desired solution included
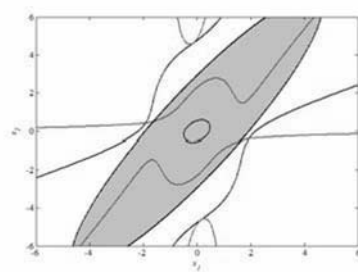
b) Dummy solution excluded

Fig. 3: Region $D(\mathbf{x}) > 0$

$$\frac{dC_A}{dt} = \frac{F}{V}\left(C_{Af} - C_A\right) - k_0 \, exp\left(\frac{-\Delta E}{RT}\right)C_A$$

$$\frac{dT}{dt} = \frac{F}{V}\left(T_f - T\right) + \left(\frac{-\Delta H}{\rho C_P}\right)k_0 \, exp\left(\frac{-\Delta E}{RT}\right)C_A - \frac{UA}{V\rho C_P}\left(T - T_j\right)$$

(12)

$$N(x,y) = 325.4977x^2 + 71.4479xy + 4.2514y^2 + 69.4087x^3 + 22.9472x^2y + 2.1447xy^2 +$$
$$0.0404y^3 + 1.7325x^4 + 0.7648x^3y + 0.0591x^2y^2 0.0008xy^3 + 0.0005y^4$$

$$D(x,y) = 1 + 0.7341x + 0.0809y + 0.0877x^2 + 0.0195xy + 0.0007y^2$$

(13)

Table 1: Nomenclature and data

Symbol	Description	Value
C_A	Concentration of A in reactor	2.3589
T	Reactor temperature	368.0629
k_0	Pre-exponential factor	$9703*3600 \text{ h}^{-1}$
$-\Delta H$	Reaction heat	5960 kcal/kgmol
ΔE	Activation energy	11843 kcal/kgmol
ρCp	(Density)(heat capacity)	500 kcal/(m^3°C)
UA/V	(Hat transfer coefficient)(Area)/(Reactor Volume)	150 kcal/(m^3°C h)
R	Ideal gas constant	1.987 kcal/(kgmol K)
F/V	Residence time	1 h^{-1}
C_{Af}	Feed Concentration	10 kgmol/m^3
T_f	Feed Temperature	298 K
T_j	Jacket temperature	298 K

The problem was solved in the GAMS platform with the global optimization solver BARON (GAMS, 2008). The level sets of the Lyapunov function and its time derivative are shown in Fig. 4. The obtained estimation is the closed curve in blue around the equilibrium. $dV(\mathbf{x})/dt=0$ is shown in red and region D($\mathbf{x}$) > 0 in gray.
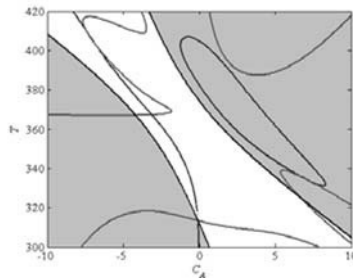
Fig. 4: Level sets $V(\mathbf{x}) = 639.5575$ and $dV(\mathbf{x})/dt = 0$ for system (12)/(13)

5. Conclusions

A global optimization formulation was proposed to estimate DOAs of stable equilibriums of systems of dynamic equations based on Lyapunov functions. In order to avoid potential dummy solutions, a tangency condition and a constraint on the denominator of $V(\mathbf{x})$ were added to previously published global optimization approaches. The rationale behind the proposed methodology was qualitatively described by means of a very rich example and also applied to a classic chemical engineering system.

References

GAMS (2008). A Users' Guide /The Solvers Manual.GAMS Development Corporation.

Hachicho, O. (2007). A novel LMI-based optimization algorithm for the guaranteed estimation of the domain of attraction using rational Lyapunov functions. Journal of the Franklin Institute. 344, pp. 535-552.

Vannelli, A. and M. Vidyasagar (1985). Maximal Lyapunov Functions and Domains of Attraction for Autonomous Nonlinear Systems. Automatica, 21, 69-80.

Probabilistic Modelling and Stochastic Dynamic Optimization for Managing Abnormal Situations in Plant-Wide Operations

Yu Yang, Jong Min Lee

Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, T6G 2G6, Canada

Abstract

This study proposes a novel dynamic real-time optimization framework that detects abnormalities at their onset and calculates a new operating condition to contain their further development. The proposed approach utilizes operational data and constructs a probability transition matrix that describes transition probabilities among representative operating modes under uncertainties. A quantitative measure of risk is defined to strike a balance between risk and profit in decision making. An integrated plant example consisting of a reactor, a storage tank, and a separator with a recycle loop is presented to demonstrate the efficacy of the approach.

Keywords: self organizing map, approximate dynamic programming, real-time optimization, plant-wide control.

1. Introduction

With an ever increasing need for improving process efficiency and product quality in today's complex manufacturing environment, real-time optimization (RTO) is considered an essential technique for the operation of complex chemical plants. Steady-state model-based RTO has been the most popular technique due to its simplicity (Cutler and Perry, 1983). However, steady-state RTO may leave significant room for performance improvement because it ignores the dynamic degrees of freedom such as unused capacity of a buffer tank and provides infeasible operating strategies during certain time periods. On the other hand, full-scale dynamic plant-wide models were used for rigorous online dynamic optimization based on a nonlinear programming method (Biegler et al., 2001). This approach may not be scalable to practical plant-wide optimization problems due to its excessive computational requirements.

One common difficulty associated with the model-based methods is obtaining a reliable plant-wide dynamic model in the form of algebraic/differential equations, apart from its relevance to online optimization. Another key issue of a model-based approach is that uncertainties from various sources can seriously degrade the performance. For example, model uncertainty can lead to RTO solutions that are far from the true plant optimum, or even an infeasible solution with respect to actual process constraints. Furthermore, once such solutions implemented and early symptoms go unnoticed, abnormal situations may arise and lead to significant periods of off-spec products or plant shutdowns.

Dynamic programming (DP) offers a unified approach to solving dynamic optimization problems under uncertainty (Bertsekas, 2005). The DP algorithm decomposes a

stochastic dynamic optimization problem into a sequence of single-period sub-problems that are solved recursively backward in time. Central to the methodology is the "value" function, which is obtained via solving Bellman's equation. Since DP algorithms compute and store a table consisting of one value per state, the size of a state space typically grows exponentially in the number of state variables, referred to as the curse of dimensionality.

Recently, the approach of approximate dynamic programming (ADP) has received much attention in the community of operations research (Si et al., 2004) and has been introduced to process systems engineering field (Lee and Lee, 2006). In most ADP methods, the value table is constructed only for a small subset of states and a function approximator is used to estimate values for unvisited states. The sample points are determined by running simulations with some known or random policies. The goal of this contribution is to develop a novel ADP-based RTO framework suitable for a large-scale plant under uncertainty.

2. Background

2.1. Markov Decision Processes and Dynamic Programming

Markov decision process (MDP) is a general modeling framework for multistage optimal control problems under uncertainty. At each of a series of discrete time steps, the state of a system (x_t) is observed and used to select an action (u_t), which then causes the system to change state and emit a numerical reward, $r(x_t, u_t)$. In general, the system dynamics is probabilistic for systems under uncertainty:

$$p_t(x_{t+1}|x_t, u_t) \tag{1}$$

which gives the probability of being in state x_{t+1} at time $t+1$ when the system is in state x_t and the decision u_t is implemented at time t.

The objective is to find a policy that maximizes the total reward received. The trade-off between immediate and delayed reward is handled by a discount factor $\gamma < 1$. Then the *value* of following a policy π from a state x is defined as the expectation of the sum of the subsequent rewards, each discounted geometrically by its delay:

$$V^\pi(x) = E\left\{\sum_{t=0}^{\infty} \gamma^t r(x_t, u_t) \middle| x = x_0\right\} \tag{2}$$

An optimal policy, π^*, achieves $V^{\pi^*}(s) \geq V^\pi(s) \; \forall \; \pi, s$. DP involves iteratively solving the following Bellman equation:

$$V^*(x_t) = \max_{u_t} E\left\{r(x_t, u_t) + \gamma V^*(x_{t+1})\right\} \tag{3}$$

Once V^* is obtained, the optimal decision is easy to find through the following single-stage optimization

$$u_t^* = \arg\max_{u_t} E\left\{r(x_t, u_t) + \gamma V^*(x_{t+1})\right\} \tag{4}$$

2.2. Approximate Dynamic Programming

In ADP, the optimal (or *nearly optimal*) value function is obtained by iteratively improving the initial estimate of optimal value function based on a set of sampled state points visited by simulation or real-operation. One of the simplest methods in ADP is temporal difference learning of order 0 (TD(0)), which simply performs an update as follows:

$$\tilde{V}(x_t) \leftarrow r(x_t, u_t) + \gamma \tilde{V}(x_{t+1}) \tag{5}$$

where $\tilde{V}$ is approximation of the optimal value function.

Another advantage of ADP in solving the Bellman equation is that it can be extended to the case where transition probability or process model is unknown. This is possible by encoding the optimal value function with state and action pair as follows:

$$Q^*(x_t, u_t) = E\left\{ r(x_t, u_t) + \gamma \max_{u_{t+1}} Q^*(x_{t+1}, u_{t+1}) \right\} \tag{6}$$

The optimal value function and the optimal action-value function have the relationship of $V^*(x) = \max_u Q^*(x, u)$. The optimal decision given x_t is simply obtained by solving

$$u_t^* = \arg \max_{u_t} Q^*(x_t, u_t) \tag{7}$$

3. Proposed RTO Framework

3.1. Construction of Markov Chain Model
The first step in the proposed approach is to model underlying behavior of plant-wide dynamics under uncertainties as a series of "representative" discrete states evolving over time without any predetermined equations. In order to characterize such representative states, we cluster process data using self-organizing map (SOM) (Kohonen, 2001). SOM creates an ordered set of states in one or more dimensions. This offers topological structure of data, which means data points that are closer to each other in high dimensional space will be projected to the neighboring clusters (or nodes) in low dimensional space.

Once the historical data set is divided into representative nodes, the probability transition matrix P and one step reward are computed. It is assumed that each unit can take a finite number of set-point values. Under a particular action vector u (i.e., a particular set of set points), the transition probability from a discrete node i to j is estimated as

$$P_{ij}^u = \frac{N_{ij}^u}{\sum_{k=1}^{n_T} N_{ik}^u} \tag{8}$$

where N_{ij}^u is the number of switches from the states belonging to the node i to those belonging to j under action u and n_T is the total number of discrete nodes. The one step reward R_{ij}^u, the reward incurred during the transition from i to j under action u, is estimated as

$$R_{ij}^u = \frac{\sum_{x_{ij}} r(x_{ij}, u)}{M} \tag{9}$$

where x is the state point that made a transition from node i to j under action u and M is the total number of transitions.

A special node, F, is also designed to represent "failure states" such as shutdown, reaction termination, or constraint violation. This node behaves like an absorbing state,

meaning that the system stays at the same state with probability of one. A negative one step reward should be assigned to this node due to the loss of performance.

Once the representative states and their transition probabilities are obtained, the optimal value for each cluster is obtained by iteratively solving the Bellman equation:

$$V^*(i) = \max_u E\left\{R_{ij}(i,u) + \gamma V^*(j)\right\} \tag{10}$$

3.2. Model Refinement

A group of data points with distinctive values from the majority of data points inside each node can occur, which will degrade the performance of RTO. In order to identify such "outliers," Q value of each process state and action pair is first estimated using

$$Q(x_t, u_t) = r(x_t, u_t) + \gamma V^*(j) \tag{11}$$

where j is a representative node containing x_{t+1}. The outliers are defined as the data points with Q values falling outside $[Q_1-1.5(Q_3-Q_1), Q_3+1.5(Q_3-Q_1)]$ (Tukey, 1977). Q_1 and Q_3 denote the first and the third quartiles, respectively.

As the outliers are detected successively in each node, simple hyper-spheres are constructed to represent the area of outliers for fast classification in real-time applications. The center of each hyper-sphere is the current outlier, and the radius is the distance between that outlier data and its nearest "normal" data. Since the probabilistic model is not valid for outliers any more, a model-free approach is employed to estimate a new value for each outlier. Once the radius and centroid of hyper-sphere is defined, the data points belonging to the area are counted and their action-values $Q(x, u)$ are computed by the following model-free update scheme:

$$Q(x_t, u_t) \leftarrow Q(x_t, u_t) + \alpha\left\{r(x_t, u_t) + \gamma Q(x_{t+1}, u_{t+1}) - Q(x_t, u_t)\right\} \tag{12}$$

where α is a step-size parameter less than 1. The model refinement step can be performed both off-line (i.e., using the historical data only) and online (i.e., when the optimizer is put to work). The transition matrix, optimal value for each node, and action-values for outliers are updated online as a new observation is available at each time step.

3.3. Quantitative Risk Term

The proposed method designs a special node, F, to represent failure states and assigns a negative value V_F. The policy derived from the learned value and action-value functions with V_F will strike a balance between total reward and the risk of entering undesirable states. In this work, V_F is assigned by

$$V_F = \frac{(D+1)\underline{r} - D\bar{r}}{\gamma(1-\gamma)} \tag{13}$$

where D is a user-specified parameter, $\underline{r}$ and $\bar{r}$ are lower and upper bounds of single-stage reward, respectively. The detailed discussion on how to derive Eq. (12) is provided in Yang and Lee (2009).

4. Case Study

An integrated process studied in Tosukhowong et al. (2004) is considered to illustrate the efficacy of the proposed framework. The process is composed of a CSTR, a storage tank, and a flash tank with a material recycle stream as shown in Fig. 1. A fresh feed

stream F_0 consisting of pure component 1 is fed to the reactor, where two irreversible reactions take place to produce the desired product 2 and undesired product 3. Detailed model descriptions can be found in Tosukhowong et al. (2004).
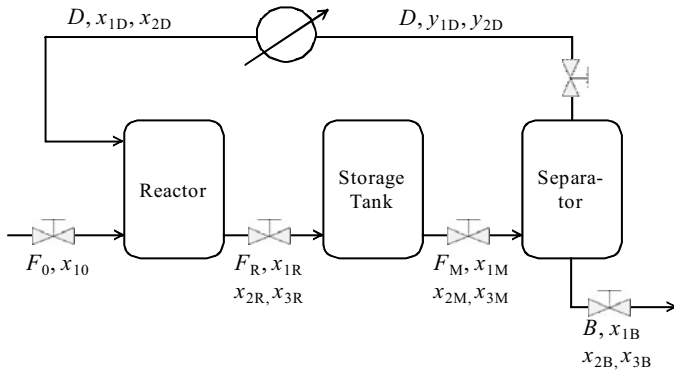
Figure 1. Schematic of the reaction-storage-separation network.

Three proportional-only (P) controllers are designed to control liquid level of each unit, and inappropriate assignment of a set point trajectory to each unit may lead to instability or significantly reduced performance under uncertainty in the inlet flow and kinetic parameters. Use of more sophisticated local controllers than P-controller may enhance the plant-wide performance. However, the question here is how to improve the plant-wide performance given the existing local controllers by coordinating operating conditions of each unit.

The objective of plant-wide optimization is to maximize the production of desired product formulated as

$$\max \sum_{t=1}^{\infty} 0.9^{t-1} B(t) x_{2B}(t) \tag{14}$$

The time interval of dynamic optimization is set as 40 minutes as suggested in Tosukhowong et al. (2004). In each episode of simulation, a sequence of random set points was used for each unit for six hours. Every 100-200 minutes, F_0 and the kinetic parameters were alternating between upper and lower bounds ($1.45 \leq F_0 \leq 1.85$, $0.01 \leq k \leq 0.03$) with their nominal values 1.667 and 0.0167, respectively.

From the data, the SOM structure with 7×3 was selected considering the number of outliers. With the 21 clusters, the single-stage reward and the transition matrices were calculated. V_F was estimated as -1478 with γ and D set as 0.9 and 4, respectively.

For comparison with steady state RTO and dynamic RTO proposed in Tosukhowong et al. (2004), 300 episodes of simulation scenarios different from the training ones were executed. As inappropriate set points will lead to total drainage of storage tank and plant shutdown, the ratio of failing optimization are compared in Fig. 2. Fig. 3 shows one of the sample realizations in the test. The production of x_{1B} is interrupted (a negative value) at 250th sample time under the dynamic RTO.

5. Conclusions

The proposed approach is based on a Markov chain model and approximate dynamic programming and provides a more robust and risk-averse policy compared to other

deterministic model-based approaches. This method can be applied to any dynamic optimization problems when a process model is difficult to obtain, which is likely for large-scale problems, and uncertainties affect operations significantly.
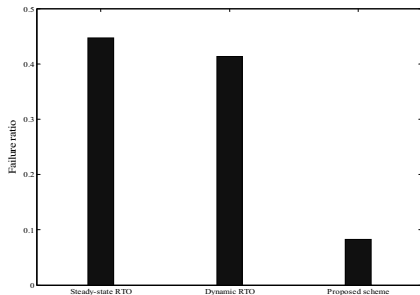
Figure 2. Ratio of operation leading to total drainage of the storage tank.
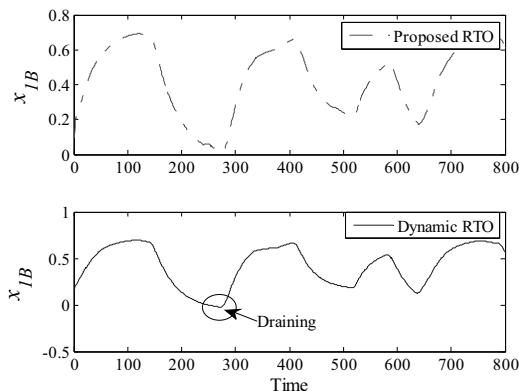
Figure 3. Constraint violation (total drainage) based on the dynamic RTO.

References

D.P. Bertsekas, 2005, Dynamic Programming and Optimal Control, 3rd Ed., Vol. 1, Athena Scientific.

C.R. Cutler, Perry, R.T., 1983, Real time optimization with multivariable control is required to maximize profits. Computers and Chemical Engineering, 7, 663-667.

T. Kohonen, 2001, Self-Organizing Maps, Springer.

J.H. Lee, J. M. Lee, 2006, Approximate dynamic programming based approach to process control and scheduling. Computers and Chemical Engineering, 30, 1603-1618.

J. Si, A.G. Barto, W.B. Powell, 2004, Handbook of Learning and Approximate Dynamic Programming, Wiley-IEEE.

T. Tosukhowong, J.M. Lee, J.H. Lee, J. Lu, 2004, An introduction to a dynamic plant-wide optimization strategy for an integrated plant. Computers and Chemical Engineering, 29, 199-208.

J.W. Tukey, 1977, Exploratory Data Analysis, Addison-Wesley.

Y. Yang, J.M. Lee, 2009, Probabilistic modeling and dynamic optimization for performance improvement and risk management of plant-wide operation, Computers and Chemical Engineering, submitted.

1293

Multi-Scenario-Based Robust Nonlinear Model Predictive Control with First Principle Models

Rui Huang[a], Sachin C. Patwardhan[b], Lorenz T. Biegler[a]

[a] *Department of Chemical Engineering, CMU, Pittsburgh, PA, USA 15213.*
[b] *Department of Chemical Engineering, IIT Bombay, Mumbai, India, 400076.*

Abstract

A robust nonlinear model predictive control (NMPC) algorithm is developed based on a multi-scenario formulation. The uncertainties are characterized by different scenarios so that the calculated control action is feasible over the entire uncertainty region. We show that this multi-scenario formulation is Input-to-State practically Stable (ISpS), and can be easily extended to the recently proposed advanced-step NMPC (as-NMPC), which is able to reduce the online computational delay. We demonstrate the advantages of this strategy on a large-scale air separation unit.

Keywords: NMPC, Robust strategy, large-scale applications

1. Introduction

Nonlinear model predictive control (NMPC) has recently received a great amount of attention, especially for energy and chemical processes that require dynamic real-time optimization (D-RTO). Moreover, a recent vehicle for D-RTO deals with advanced-step Nonlinear Model Predictive Control (as-NMPC) (Zavala & Biegler, 2008), which incorporates very large dynamic optimization problems and can reduce the on-line feedback delay to only a few CPU seconds, even with industrial-size first principle dynamic models.

On the other hand, ensuring stability of closed-loop systems with faults and model mismatch remains an important problem in NMPC. While nominal stability and feasibility have been extensively addressed in the past decades (Magni & Scattolini, 2007, Limon et al, 2008), NMPC with asymptotic stability does not guarantee robust stability and large unmeasured disturbances can lead to poor performance for D-RTO. In these cases, robust design strategies are necessary to account for uncertainties explicitly in the controller formulation. A well known strategy to guarantee the robust stability is the min-max NMPC formulation, which computes the best control policy based on the worst expected realization of the uncertainties. However this formulation dramatically increases the computational cost of the on-line NMPC problem, despite some recent remedies (Diehl et al, 2008, Fontes & Magni, 2003). Moreover, since the control action is based on the worst case, the performance is compromised, e.g. output variables present large offset.

In this work, we propose a robust NMPC design strategy based on multi-scenario NLP formulation and extend it to as-NMPC. The robust stability can be established through the recent input-to-state practical stability (ISpS) framework (Limon et al, 2008). Section 2 presents the multi-scenario formulation for NMPC and as-NMPC and analyze their robust stabilities; section 3 shows that the proposed method is suitable for real world systems by considering a large-scale air separation unit with plant-model mismatches; section 4 concludes the paper.

2. NMPC based on Multi-scenario formulation and Stability Analysis

2.1. Multi-scenario Formulation

In this work, the dynamics of an uncertain plant will be described by the following discrete-time model,

$$x(k+1) = f(x(k), u(k), \phi(k)), \ k \geq 0 \tag{1}$$

where $x(k) \in \mathfrak{R}^{n_x}$ is a system state vector, $u(k) \in \mathfrak{R}^{n_u}$ is a control variable vector and $\phi(k) \in \mathfrak{R}^{n_\phi}$ is an uncertainty signal which models disturbances and plant-model mismatches at time steps $k \geq 0$. For the sake of clarity, we only consider the state-independent uncertainty signals in this work, i.e. $\phi(k)$ does not depend on $x(k)$ and $u(k)$. However with slight modification of notations, the analysis here can be extended to the case with state-dependent uncertainties. In general, the control and state of the plant are required to fulfill the constraints, e.g. $x(k) \in \mathbb{X}$, $u(k) \in \mathbb{U}$, and only the partial information of the uncertainty $\phi(k) \in \Omega$ is available, e.g. its feasible region Ω. Without losing generality, we assume that the given plant (1) has an equilibrium point at the origin, i.e. $f(0, 0.0) = 0$.

Given $x(k)$, the current state value at time step t_k, the NMPC formulation can be described in the following discretized form:

$$\min_{z(j), v(j)} \ J(x(k)) := F(z(N)) + \sum_{j=0}^{N-1} \varphi(z(j), v(j), \theta(j)) \tag{2a}$$

$$s.t. \ z(j+1) = f(z(j), v(j), \theta(j)), \ z(0) = x(k) \tag{2b}$$

$$z(j) \in \mathbb{X}, \ z(N) \in \mathbb{X}_f, \ v(j) \in \mathbb{U}, \ \theta(j) \in \Omega, \ j = 0, \dots N - 1 \tag{2c}$$

where N is the finite time horizon, $\theta \in \Omega$ is the uncertainty parameter in the controller. The computed control $v(j) \in \mathfrak{R}^{n_u}$ and predicted state $z(j) \in \mathfrak{R}^{n_x}$ are enforced to satisfy the constraints $v(j) \in \mathbb{U}$, $z(j) \in \mathbb{X}$, $z(N) \in \mathbb{X}_f$. The cost function $J(x(k))$ at t_k comprises stage cost $\varphi(\cdot, \cdot, \cdot)$ and terminal penalty cost $F(\cdot)$. When the solution sequence $(z^*(l), v^*(l))$ is available, the first control $u(k) = v^*(0)$ is injected into the plant. In general, $x(k+1)$ is different from the predicted value because of the plant-model mismatches, (e.g. $\theta \neq \phi$). At the next time step t_{k+1} when $x(k+1)$ is measured or estimated, we move the time horizon one step further and solve problem (2) again.

If the presence of uncertainties does not cause any loss of feasibility (e.g. no constraints), the NMPC (2) enjoys inherent robustness (Magni & Scattolini, 2007). However if the constraints are enforced and uncertainties are present, the calculated control v from (2) may be infeasible for the plant (1). To avoid this difficulty, we solve a multi-scenario problem:

$$\min_{z_l(j), v(j)} \ V(x(k)) := \sum_{l=1}^{M} w_l J_l(x(k))$$

$$= \sum_{l=1}^{M} w_l \left\{ F_l(z_l(N)) + \sum_{j=0}^{N-1} \varphi(z_l(j), v(j), \theta_l) \right\} \tag{3a}$$

$$s.t. \ z_l(j+1) = f(z_l(j), v(j), \theta_l), \ z_l(0) = x(k), \ l = 1, \dots M \tag{3b}$$

$$z_l(j) \in \mathbb{X}, \ z_l(N) \in \mathbb{X}_f, \ v(j) \in \mathbb{U}, \ \theta_l \in \Omega, \ j = 0, \dots N - 1, \tag{3c}$$

with M different uncertainties $\theta_l, l \in 1, \dots, M$, where l is the index of scenarios and w_l are weights for each scenario, satisfying $0 \leq w_l \leq 1$, $\sum_{l=1}^{M} w_l = 1$. In addition, consider $B_{\delta l}$ as M closed balls, $B_{\delta l} \triangleq \{\vartheta_l \in \Omega | |\vartheta_l - \theta_l| \leq \delta_l\}$, $\forall l \in 1 \dots M$, centered around θ_l with radius δ_l. The balls $B_{\delta l}$ are defined such that an NMPC formulated with

θ_l as nominal model parameter is robustly stable within this ball. We will prove that the control sequence $v(\cdot)$ calculated from (3) remains feasible and stable for all the uncertainty regions $B_{\delta_l}, \forall l$. If $\phi(j) \in \bigcup_{l=1}^M B_{\delta l}, \forall j$, then the closed-loop system is robustly stable.

2.2. Robust Stability Analysis

For the analysis of the robust stability of multi-scenario NMPC, we make use of the regional input-to-state practical stability (ISpS) property developed in Limon et al, 2008. Here, we only summarize key definitions and results.

Definition 1. *Given a set $\Gamma \subseteq \Re^{n_x}$, including the origin as an interior point, the system (1) is said to be ISpS in Γ w.r.t ϕ, if Γ is a robust positive invariant set for (1) and there exist a $\mathcal{KL}$ function β, a $\mathcal{K}$ function γ and a constant $c \geq 0$ such that,*

$$|x(k)| \leq \beta(|x_0|, k) + \gamma(|\phi_{(k-1)}|) + c, k \geq 0, \ \forall x_0 \in \Gamma, \ \phi(k) \in \Omega \qquad (4)$$

with $|\phi_{(k-1)}| \triangleq sup_{k\geq 0}\{|\phi(k-1)|\}$. Moreover, a function $V(\cdot)$ is an ISpS-Lyapunov function for system (1) in Γ if there exist $\mathcal{K}$ functions α_1, α_2, α_3, σ and constants $c_1, c_2 \geq 0$ such that:

$$\alpha_2(|x|) \leq V(x) \leq \alpha_2(|x|) + c_1, \ \forall x \in \Gamma$$
$$\triangle V(x, \phi) = V(f(x, u, \phi)) - V(x) \leq -\alpha_3(|x|) + \sigma(|\phi|) + c_2. \forall x \in \Gamma, \ \forall \phi \in \Omega \qquad (5)$$

Lemma 1. *If (1) admits an ISpS-Lyapunov function in Γ, then it is ISpS in Γ w.r.t ϕ.*

In order to deal with the multi-scenario formulation, the common robust stability assumptions (Magni & Scattolini, 2007, Zavala & Biegler 2008, Limon et al, 2008) are modified as follows.

Assumption 1 *(Assumptions for Multi-scenario Formulation and Robust Stability)*
a) There exists a constant $c_3 \geq 0$, that the terminal penalty $F_l(\cdot)$ satisfy $\alpha_F(|x(k)|) \leq F_l(x(k)) \leq \gamma_F(|x(k)|) + c_3, \ \forall l, \ \forall x \in \mathbb{X}_f \subseteq \Gamma, \ \alpha_F(\cdot)$ and $\gamma_F(\cdot)$ are $\mathcal{K}$ functions.
b) There exists a local control law $h_f(z_l(N))$ defined on $\mathbb{X}_f \subseteq \Gamma$, such that $f(z_l(N), h_f(z_l(N)), \theta_l) \in \mathbb{X}_f, \forall z_l(N) \in \mathbb{X}_f,$ and $F_l(f(z_l(N), h_f(z_l(N)), \theta_l)) - F_l(z_l(N)) \leq -\varphi(z_l(N), h_f(z_l(N)), \theta_l), \ \forall l.$
c) $\alpha_p(|x(k)|) \leq \varphi(x(k), u(k), \theta(k)) \leq \alpha_q(|x(k)|) + c_4, \forall \theta(k) \in \Omega,$ for some constant $c_4 \geq 0$ with $\alpha_p(\cdot), \alpha_q(\cdot)$ as $\mathcal{K}$ functions.
d) Given $u(k)$ and $z_l(k+1) = f(x(k), u(k), \theta_l)$ for each scenario l, there exist future mismatches at the next sampling time:

$$\epsilon_l(x(k+1)) := J_l(x(k+1)) - J_l(z_l(k+1)) \ \forall l \qquad (6)$$

with $\mathcal{K}$ functions α_4, σ_2 and positive Lipschitz constants L_c such that $|\epsilon_l(x(k+1))| \leq L_c(\alpha_4(|x(k)|) + \sigma_2(|\theta_l - \phi(k)|)).$
e) For a constant $c_5 \geq 0$, $\mathcal{K}$ functions α_5 and $\theta_l \in \Omega, \ \forall l,$ there exists a constant $K \geq 0$ such that:

$$\sum_{l=1}^M w_l\left[-\varphi(x(k), u(k), \theta_l) + K(\alpha_4(|x(k)|) + \sigma_2(|\theta_l - \phi(k)|))\right]$$
$$\leq -\alpha_5(|x(k)|) + \sigma_1(|\phi(k)|) + c_5 \qquad (7)$$

Robust stability of multi-scenario NMPC can now be established by the following theorem.

Theorem 1 *(Robust ISpS stability of Multi-scenario NMPC) If the plant uncertainty parameter vector lies in the union of balls centered around the uncertainty parameters in the controller, i.e. $\phi(j) \in \bigcup_{l=1}^{M} B_{\delta l}, \forall j$, then under Assumption 1, with $K \geq L_c$, the cost function $V(x(k)) = \sum_{l=1}^{M} w_l \left[F_l(z_l(N)) + \sum_{j=0}^{N-1} \varphi(z_l(j), v(j), \theta_l) \right]$ is an ISpS-Lyapunov function for plant (1) in Γ and the resulting closed loop system is ISpS stable.*

Proof. First let $u(k)$ be the first element of the control sequence calculated from (3) based on $x(k)$ at t_k. There exists a $\theta_{min}(k) \in \Omega$ so that $\varphi(x(k), u(k), \theta_l) \geq \varphi(x(k), u(k), \theta_{min}(k)), \forall \theta_l \in \Omega$ for the fixed $x(k)$ and $u(k)$, then

$$V(x(k)) = \sum_{l=1}^{M} w_l \left\{ F_l(z_l(N)) + \sum_{j=0}^{N-1} \varphi(z_l(j), v(j), \theta_l) \right\}$$

$$\geq \sum_{l=1}^{M} w_l \sum_{j=0}^{N-1} \varphi(z_l(j), v(j), \theta_l)$$

$$\geq \sum_{l=1}^{M} w_l \varphi(x(k), u(k), \theta_{min}(k)) \geq \alpha_p(|x(k)|) \tag{8}$$

the last two inequalities follow from letting $j = 0$ and Assumption 1.c).

Then we consider the multi-scenario NMPC with horizon N+1, its objective function is

$$V(x(k), N+1) = V(x(k), N) + \sum_{l=1}^{M} w_l \Big[-F_l(z_l(N)) +$$

$$F_l(f(z_l(N), h_f(z_l(N)), \theta_l)) + \varphi(z_l(N), h_f(z_l(N)), \theta_l) \Big] \leq V(x(k), N) \tag{9}$$

the inequality is from Assumption 1.b). So that

$$V(x(k), N+1) \leq V(x(k), N) \leq \ldots \leq V(x(k), 0)$$

$$= \sum_{l=1}^{M} w_l F_l(x(k)) \leq \gamma_F(|x(k)|) + c_3 \tag{10}$$

the last inequality is from Assumption 1.a).

Finally, we are able to compare the cost function $V(x(k+1))$ and $V(x(k))$ from the two neighboring problems with horizon length N,

$$V(x(k+1)) - V(x(k))$$

$$= \sum_{l=1}^{M} [w_l J_l(z_l(k+1))] - V(x(k)) + V(x(k+1)) - \sum_{l=1}^{M} [w_l J_l(z_l(k+1))]$$

$$= \sum_{l=1}^{M} w_l [J_l(z_l(k+1)) - J_l(x(k))] + \sum_{l=1}^{M} w_l \epsilon_l(x(k+1)) \tag{11}$$

From Assumption 1.b), we know $J_l(z_l(k+1)) - J_l(x(k)) = F_l(f(z_l(N), h_f(z_l(N)), \theta_l)) - F_l(z_l(N)) + \varphi(z_l(N), h_f(z_l(N)), \theta_l) - \varphi(x(k), u(k), \theta_l) \leq -\varphi(x(k), u(k), \theta_l), \forall l$. Consequently

$$V(x(k+1)) - V(x(k)) \leq \sum_{l=1}^{M} w_l [-\varphi(x(k), u(k), \theta_l) + \epsilon_l(x(k+1))]$$

$$\leq \sum_{l=1}^{M} w_l \Big[-\varphi(x(k), u(k), \theta_l) + L_\epsilon(\alpha_4(|x(k)|) + \sigma_2(|\theta_l - \phi(k)|)) \Big] \tag{12}$$

where (12) follows from Assumption 1.d). With $K \geq L_c$ and Assumption 1.e),

$$V(x(k+1)) - V(x(k)) \leq -\alpha_5(|x(k)|) + \sigma_1(|\phi(k)|) + c_5 \tag{13}$$

The desired result follows as (8), (10) and (13) allow *x(k)* and *V* to satisfy (4) and (5). ?

Note from the analysis, the robust stability is guaranteed when $\phi(j) \in \bigcup_{l=1}^{M} B_{\delta l}, \forall j$, we also see that the it will improve if the majority of the uncertainty parameters in the controller are close to that in the plant. In practice, since the plant uncertainty parameter is unknown, the multi-scenario NMPC can be initiated with a few guessed parameters, if the controller appears to be unstable, more scenarios should be added to make sure that the plant uncertainty ϕ is in the neighborhood of one uncertainty parameter θ_l, hence the robust stability is achieved.

2.3. Multi-scenario for as-NMPC

Although the multi-scenario formulation can improve robust stability, it may introduce large computational delay online. In order to reduce the online computational delay, we modify the algorithm of the advanced-step NMPC (Zavala & Biegler, 2008) to incorporate the multi-scenario formulation, as follows:

1. Background calculation: at t_k, predict the future state $z_l(k+1)$ with $x(k)$ and $u(k)$ for each scenario l. Solve problem (3) with $z_l(k+1)$ as the initial condition for each scenario l. The NLP solver takes a Newton step, which can be expressed as:

$$K(z_l(k+1))\triangle s = -\varphi(z_l(k+1)) \tag{14}$$

where s is the solution vector and $\triangle s$ represents the Newton step from the current solution towards the new point, $K(\cdot)$ and $\varphi(\cdot)$ are the KKT matrix and KKT condition at current solution, respectively.

2. Online update: at t_{k+1}, obtain $x(k+1)$ and use (14) to get the updated solution: $K(z_l(k+1))\triangle s = -\varphi(x(k+1))$. Extract control action $u(k+1)$ and inject into plant.

3. Iterate: set $k \leftarrow k+1$, and go to step 1.

Note that the KKT matrix in (14) has already been factorized in solving (14), so only a single back-solve is required in Step 2; this is at least two orders of magnitude faster than solving the optimization problem (3). Along the same lines, we can derive a similar but weaker ISpS property for multi-scenario as-NMPC, where the online update leads to only a small deterioration of the stability property.

3. Simulation examples

We now revisit the large-scale air separation unit reported in Huang et al, 2009 with plant-model mismatches. In this work, we modify the ASU model to introduce a tray efficiency, which is considered as an uncertain parameter. In this simulation, we assume that the true value of tray efficiency is $\alpha = 0.8$ and the ASU starts from a nominal steady-state. The oxygen (POX) and nitrogen production (PNI) rate set-points are reduced by *30%* through a ramp change from $t = 30$ to $t = 60$ minutes, while temperatures (Tl30 & Th15) are required to remain constant to control the product purities. After this, they undergo a ramp increase from $t = 1000$ to $t = 1030$ back to their original values. The NMPC formulation with $\alpha = 0.9$ or $\alpha = 1$, and multi-scenario NMPC with two scenarios $\alpha = 0.9$ and $\alpha = 1$ yield infeasible control actions. Hence, we choose equally weighted three scenarios with $\alpha_1 = 0.8$, $\alpha_2 = 0.9$ and $\alpha_3 = 1$ in the multi-scenario as-NMPC. Figure 1 shows that it is able to track the output variables with stable control actions and product purities are satisfied. Moreover multi-scenario as-NMPC yields nearly the same performance as the ideal multi-scenario NMPC. Without plotting results, the same NMPC formulation is robustly stable for cases where tray efficiency equals to 0.85, 0.95 and 1. After discretization, the optimization problem with three scenarios leads to a NLP contains 350,940 variables and 350,700 constraints. The simulation was done on an Intel Quad Core @ 2.4GHz server running Linux. For each horizon, IPOPT took 20-30 CPU minutes for the background calculation, while

sensitivity update took around 2 CPU seconds. Note online computation time is reduced by at least 600 times with virtually no loss in performance.
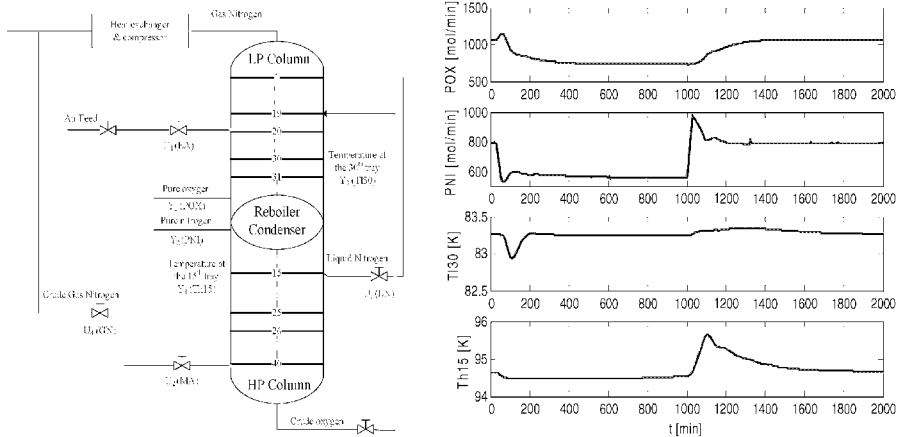
Figure 1. ASU flowsheet and output variable profiles. Dash-dotted- lines are setpoints, superimposed dashed and solid lines are multi-scenario NMPC and as-NMPC, respectively.

4. Concluding Remarks

In this study a robust NMPC strategy based on multi-scenario formulation is presented. This strategy solves multiple problems with different uncertainty parameters to guarantee the satisfaction of constraints. We establish the input-to-state practical stability of this formulation for both NMPC and as-NMPC. The advantages of the proposed method are demonstrated by simulation of a large-scale ASU. In addition the multi-scenario as-NMPC presents nearly the same performance as the NMPC counterpart, but dramatically reduces the online computation delay. For the future work, we propose to study the strategies that allow the background solution taking more than one sampling time while still providing instantaneous on-line update. Besides parallel computing with distributed computer cluster will be studied to future reduce the background computational effort. Moreover we will explore methods to reduce the number of scenarios needed to guarantee the robust stability, e.g. back-off constraints.

References

M. Diehl, J. Gerhard, W. Marquardt, M. Mönnigmann, 2008, "Numerical solution approaches for robust nonlinear optimal control problems", Computers & Chemical Engineering, 32: 1279.

F. Fontes & L. Magni, 2003, "Min-max model predictive control of nonlienar systems using discontinuous feedbacks", IEEE Transactions on Automatic Control, 48: 1750.

R. Huang, V. Zavala & L. Biegler, 2009, "Advanced step noninear model predictive control for air separation units", J. of Process Control, 19: 678

D. Limon, T. Alamo, D. Raimondo, D.M. Peña, J. Bravo, E. Camacho, 2008. "Input-to-state stability: an unifying framework for robust model predictive control", In Assessment and Future Directions of Nonlinear Model Predictive Control, Springer, in press

L. Magni & R. Scattolini, 2007, "Robustness and robust design of mpc for nonlinear discrete-time systems", Assessment and Future Directions of Nonlinear Model Predictive Control, 239-254.

V. Zavala & L. Biegler, 2008, "The advanced step nmpc controller: Optimality, stability and robustness", Automatica, 45: 86.

Approximation of Arrival Cost in Moving Horizon Estimation Using a Constrained Particle Filter

Rodrigo López-Negrete[a], Sachin C. Patwardhan[b], Lorenz T. Biegler[a]

[a] *Department of Chemical Engineering, CMU, Pittsburgh, PA, USA 15213.*
[b] *Department of Chemical Engineering, IIT Bombay, Mumbai, India, 400076.*

Abstract

Moving horizon estimation (MHE) is increasingly being used for state estimation in model predictive control schemes. Approximation of the arrival cost in MHE formulation is an open issue in this domain. In this work, we propose to use a constrained particle filter for estimating the moments of the prior probability density function of the initial state and approximating the arrival cost. The efficacy of the proposed approach is illustrated in a simulation case study on a benchmark problem.

Keywords: Moving Horizon Estimation, Arrival Cost, Constrained Ensemble Kalman Filter

1. Introduction

Moving horizon estimation is a preferred approach for state estimation particularly when used in connection with model predictive control schemes (Zavala and Biegler, 2008). The main advantage of MHE is that it can handle bounds on state and parameter estimates or irregular measurements in a systematic manner. Moreover, it provides a unifying framework for state estimation in systems described by differential algebraic equations or hybrid systems. Recent focus of research in this area has been on development of computationally efficient algorithms for on-line implementation (Zavala and Biegler, 2008). An important aspect of MHE formulation is estimation of the arrival cost. In a recent review paper, Bakshi and Rawlings (2006) have indicated that best choice of the arrival cost remains an open issue in the domain of MHE research. As the true arrival cost is difficult to determine, it is often estimated using linearization approximation as in the extended Kalman filtering (EKF) formulation. This method of approximating the arrival cost, however, does not consider the bounds on state estimates in a systematic manner.

Recently, a new class of recursive Bayesian filtering technique, called particle filtering, has attracted attention of many researchers (Rawlings and Bakshi, 2006). The objective of the recursive Bayesian state estimation problem is to find the mean and variance of the random state variable using its probability density function conditioned on the past measurements. While dealing with nonlinear systems, the exact analytical solution to the recursive propagation of the posterior density is difficult to obtain. Thus, it becomes necessary to develop approximate and computationally tractable sub-optimal solutions to the sequential Bayesian estimation problem. The particle filter is a numerical method for implementing an optimal recursive Bayesian filter through Monte-Carlo simulation. Thus, even when the state estimation error has multimodal and non-Gaussian distributions, a particle filter formulation can accurately estimate its moments. Ensemble Kalman filter (EnKF) developed by Evenson (2003) is a form of particle filter that retains the computational simplicity of EKF without requiring linearization or

normality of the conditional density function. However, as proposed, EnKF formulation cannot deal with bounds on states or parameters. Recently, Prakash et al. (2007) have developed constrained version of ensemble Kalman filter (C-EnKF), which provides a systematic and statistically consistent approach to deal with bounds on the state estimates. In the present work, we exploit this feature of C-EnKF for approximating the arrival cost in the MHE formulation. The efficacy of the proposed approach is illustrated by conducting simulation studies on a benchmark problem in the state estimation literature.

2. Review of MHE and C-EnKF

2.1. Preliminaries

Consider a nonlinear system represented by the following nonlinear state space equations:

$$\mathbf{x}(k) = F\big[\mathbf{x}(k-1), \mathbf{u}(k-1)\big] + \mathbf{w}(k-1) \tag{1}$$

$$\mathbf{y}(k) = H\big[\mathbf{x}(k)\big] + \mathbf{v}(k) \tag{2}$$

In the above process model, $\mathbf{x}(k) \in R^n$ is the system state vector, $\mathbf{u}(k) \in R^m$ is known system input, $\mathbf{w}(k) \in R^p$ is the state noise with known distribution, $\mathbf{y}(k) \in R^r$ is the measured state variable and $\mathbf{v}(k) \in R^r$ is the measurement noise with known distribution. The index k represents the sampling time. The operators $F[.]$ and $H[.]$ represent the nonlinear process model and nonlinear measurement model, respectively. It is further assumed that the initial state of the system $\mathbf{x}(0)$ is a random vector with known probability distribution.

2.2. Moving Horizon Estimation

In most physical systems, states/parameters are bounded, which introduces constraints on state/parameter estimates. One major limitation of recursive nonlinear filters is that these formulations cannot handle such constraints. Moving horizon estimation (MHE) (Bakshi and Rawlings, 2006) formulation provides a systematic approach to handling of bounds on states/parameters or any other algebraic constraints. The MHE problem can be stated as (Bakshi and Rawlings, 2006)

$$\underset{\mathbf{x}(k-N:k)}{Min} \Pi\big[\mathbf{x}(k-N)\big] + \sum_{i=k-N+1}^{k} \big\{\log\big[p\big(\mathbf{w}(i-1)\big)\big] + \log\big[p\big(\mathbf{v}(i)\big)\big]\big\} \tag{3}$$

$$\text{subject to model equations (1) and (2) and} \quad \mathbf{x}_L \leq \mathbf{x}(i) \leq \mathbf{x}_U \tag{4}$$

where, $p(.)$ denotes the probability density function of the argument. The term $\Pi\big[\mathbf{x}(k-N)\big]$, called the *arrival cost,* represents the negative log of the conditional density $p\big[\mathbf{x}(k-N)\,|\,\mathbf{Y}_{k-N}\big]$ where $\mathbf{Y}_{k-N} = \big\{\mathbf{y}(1),....\mathbf{y}(k-N)\big\}$. The arrival cost represents the information in the prior measurement sequence $\mathbf{Y}_{k-N}$ that is not considered in the horizon at time k (Bakshi and Rawlings, 2006). The main difficulty with the arrival cost term is that, except for the linear unconstrained case, the conditional density $p\big[\mathbf{x}(k-N)\,|\,\mathbf{Y}_{k-N}\big]$ cannot be easily characterized using closed form algebraic expressions.

In most practical situations, the probability density functions of the state noise and the measurement noise can be well approximated using Gaussian density functions. In addition, if we further assume that the *a-priori* estimate is normally distributed

$N(\hat{\mathbf{x}}(k-N),\mathbf{P}(k-N))$ such that covariance $\mathbf{P}(k-N)$ is invertible, then the MHE objective function can be reformulated as follows

$$\underset{\Omega_k}{Min}\,\Phi\big[\mathbf{x}(k-N)\big]+\sum_{i=k-N+1}^{k}\big[\mathbf{w}(i-1)^T Q^{-1}\mathbf{w}(i-1)+\mathbf{v}(i)^T \mathbf{R}^{-1}\mathbf{v}(i)\big] \tag{5}$$

$$\Omega_k \equiv \big\{\mathbf{x}(k-N),\mathbf{w}(n-N+1),...\mathbf{w}(k-1)\big\}$$

$$\Phi\big[\mathbf{x}(k-N)\big]=\varepsilon(k-N)^T \mathbf{P}(k-N)^{-1}\varepsilon(k-N)+\Phi^*\big[\mathbf{x}(k-N)\big] \tag{6}$$

$$\varepsilon(k-N)=\mathbf{x}(k-N)-\hat{\mathbf{x}}(k-N)$$

where $\mathbf{w}(k)\sim N(\overline{\mathbf{0}},\mathbf{Q})$, $\mathbf{v}(k)\sim N(\overline{\mathbf{0}},\mathbf{R})$ and $\Phi^*\big[\mathbf{x}(k-N)\big]$ represents the optimal cost at time *(k - N)*. Similar to extended Kalman filter, an approximation for the covariance $\mathbf{P}(k-N)$ can be constructed by solving the Riccati equations with local linearization of operators F[.] and H[.] (Bakshi and Rawlings, 2006; Zavala et al., 2008).

2.3. Constrained Ensemble Kalman Filter

The filter is initialized by drawing N samples $\hat{\mathbf{x}}^{(i)}(0\,|\,0)$ from the given distribution. At each time step, N samples $\big\{\mathbf{w}^{(i)}(k),\mathbf{v}^{(i)}(k):i=1,2,...N\big\}$ are drawn randomly using the distributions of the state noise and the measurement noise. These sample points together with particles $\big\{\hat{\mathbf{x}}^{(i)}(k-1\,|\,k-1):i=1,2,...N\big\}$ are then propagated through the system dynamics to obtain transformed particles as follows:

$$\mathbf{x}^{(i)}(k\,|\,k-1)=F\big[\mathbf{x}^{(i)}(k-1\,|\,k-1),\mathbf{u}(k-1)\big]+\mathbf{w}^{(i)}(k-1) \tag{7}$$

for i = 1,2,…N. The transformed sample points that violate the bounds are projected on the constraint boundary to obtain a cloud of predicted particles $\big\{\hat{\mathbf{x}}_c^{(i)}(k\,|\,k-1):i=1,2,...N\big\}$) where $\hat{\mathbf{x}}_c^{(i)}(k\,|\,k-1)=\mathbf{P}_r\big[\hat{\mathbf{x}}^{(i)}(k\,|\,k-1)\big] \tag{8}$

and $\mathbf{P}_r\big[.\big]$ represents the projection operator. The sample covariance is estimated as

$$\overline{\mathbf{x}}_c(k\,|\,k-1)=(1/N)\sum_{i=1}^{N}\hat{\mathbf{x}}_c^{(i)}(k\,|\,k-1) \tag{9}$$

$$\mathbf{P}(k\,|\,k-1)=(1/N)\sum_{i=1}^{N}\varepsilon^{(i)}(k\,|\,k-1)\big[\varepsilon^{(i)}(k\,|\,k-1)\big]^T \tag{10}$$

where $\varepsilon^{(i)}(k\,|\,k-1)=\hat{\mathbf{x}}_c^{(i)}(k\,|\,k-1)-\overline{\mathbf{x}}_c(k\,|\,k-1)$. The updated state estimates are then obtained by solving the following set of constrained optimization problems

$$\underset{\mathbf{x}(k)}{Min}\,\big[\xi^{(i)}(k)^T \mathbf{P}(k\,|\,k-1)^{-1}\xi^{(i)}(k)+\mathbf{e}^{(i)}(k)^T \mathbf{R}^{-1}\mathbf{e}^{(i)}(k)\big] \tag{11}$$

$$\xi^{(i)}(k)=\mathbf{x}(k)-\mathbf{x}_c^{(i)}(k\,|\,k-1)\ ;\ \ \mathbf{e}^{(i)}(k)=\mathbf{y}(k)-H\big[\mathbf{x}(k)+\mathbf{v}^{(i)}(k)\big]$$

subject to constraints $\mathbf{x}_L \le \mathbf{x}(k)\le \mathbf{x}_U$. The solutions of the above set of optimization problems yields updated particles $\big\{\hat{\mathbf{x}}^{(i)}(k\,|\,k):i=1,2,...N\big\}$. The updated state estimate is then computed as follows

$$\overline{\mathbf{x}}(k\,|\,k)=(1/N)\sum_{i=1}^{N}\hat{\mathbf{x}}^{(i)}(k\,|\,k) \tag{12}$$

2.4. Generation of Truncated Distribution of Initial State

When states have bounds, it becomes necessary to generate particles that are consistent with these bounds. This can be achieved by using the concept of truncated distributions. A truncated distribution is a conditional distribution that is conditioned on the bounds on the random variable. In particular, the truncated univariate normal distribution can be obtained as follows:

$$N_T[0,1 \mid a \le \zeta \le b] = \frac{1}{\sigma\sqrt{2\pi}} \frac{\exp(-\zeta^2/2)}{\phi(b) - \phi(a)} \tag{13}$$

Where $\phi(a)$ and $\phi(b)$ are the cumulative normal distribution functions evaluated at the lower and upper bounds, respectively. Now, consider a situation in which the distribution of random variable vector $\mathbf{x}$ is approximated by truncated multivariate Gaussian density function, denoted as $\mathbf{N}_T[\bar{\mathbf{x}}, \mathbf{P}]$, and defined over bounds $(\mathbf{x}_L, \mathbf{x}_H)$. In this work, we make use of the following approach for the generation of samples from the truncated normal distribution (Prakash et al., 2007). Since $\mathbf{P}$ is a symmetric and positive definite matrix, Cholesky factorization of $\mathbf{P}$ yields a lower triangular matrix $\mathbf{S} = \sqrt{\mathbf{P}}$. Now, a transformed random vector $\mathbf{T}$ is defined such that $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{S}\mathbf{T}$. With this transformation, we can now define n-truncated univariate normal distributions $N_T^{(i)}[0,1 \mid t_{L,i} < t_i \le t_{H,i}]$ where the limits of ith truncated distribution are defined as follows:

$$t_{L,i} = \left(\mathbf{x}_{L,i} - \bar{\mathbf{x}}_i - \sum_{r=1}^{i-1} s_{ir} t_r\right)/s_{ii} \quad ; \quad t_{H,i} = \left(\mathbf{x}_{H,i} - \bar{\mathbf{x}}_i - \sum_{r=1}^{i-1} s_{ir} t_r\right)/s_{ii} \tag{14}$$

where s_{ij} represents (i,j)'th element of matrix $\mathbf{S}$. The above transformation requires that we draw samples recursively. Thus, first t_1 can be drawn from $N_T^{(1)}[0,1 \mid t_{L,1} \le \zeta \le t_{H,1}]$. Then, after setting limits $(t_{L,2}, t_{H,2})$, t_2 can be drawn from $N_T[0,1 \mid t_{L,2} \le \zeta \le t_{H,2}]$, and so on. Thus, a sample from the truncated normal distribution $\mathbf{N}_T[\bar{\mathbf{x}}, \mathbf{P}]$ can be expressed as follows

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{S}T \quad ; \quad T = \begin{bmatrix} t_1 & t_2 & \dots & t_n \end{bmatrix}^T \tag{15}$$

In the Constrained EnKF formulation, the above mentioned procedure is used for generating the initial ensemble.

3. Estimation of Arrival Cost using C-EnKF

There are two conceptual difficulties associated with the arrival cost approximation described in Section 2.2.

(a) If we decide to use normal distribution to approximate the arrival cost, it is important to use truncated normal distribution in the presence of bounds on states.

(b) The covariance update step involves linearization of nonlinear operators F[.] and H[.] in equations (1) and (2), respectively. Depending on the nature of local nonlinearity, this further introduces error in the approximation of the arrival cost.

A possible remedy to the problem of approximating arrival cost can be to construct the conditional density $p[\mathbf{x}(k-N \mid \mathbf{Y}_{k-N})]$ using a particle filter formulation that systematically handles the constraints. Since particle filters use Monte Carlo simulation approach, there is no assumption involved on the distribution of the conditional density function. If we allow infinite sample size, then, in principle, we can construct an accurate approximation of the moments of the conditional probability density function even if it is multi-modal and non-Gaussian. Moreover, operators F[.] and H[.] can be directly used for estimation of the moments without requiring the linearization step.

In this work, we retain the form of arrival cost approximation given by equation (6). The second moment of $p[\mathbf{x}(k-N) \mid \mathbf{Y}_{k-N}]$ necessary for constructing this term is approximated by employing C-EnKF over interval *[k – N - 1, k - N]*. Defining posteriori estimation error as

$$\mathbf{\epsilon}^{(i)}(k-N\,|\,k-N) = \hat{\mathbf{x}}^{(i)}(k-N\,|\,k-N) - \overline{\mathbf{x}}(k-N\,|\,k-N), \qquad (16)$$

the covariance $\mathbf{P}(k-N\,|\,k-N)$ can be estimated as follows

$$\mathbf{P}(k-N\,|\,k-N) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{\epsilon}^{(i)}(k-N\,|\,k-N)\left[\mathbf{\epsilon}^{(i)}(k-N\,|\,k-N)\right]^{T} \qquad (17)$$

The arrival cost is the approximated by setting $\mathbf{P}(k-N) = \mathbf{P}(k-N\,|\,k-N)$ in equation (6). It may be noted that some modifications are necessary in C-EnKF formulation described in Section 2.3 while using it for approximation of arrival cost. In the conventional formulation, the particles $\left\{\hat{\mathbf{x}}^{(i)}(k-N\,|\,k-N):i=1,2,...N\right\}$ are propagated while performing computations over interval $[k\text{-}N,\ k\text{-}N+1]$. In the proposed formulation, however, we only propagate the cloud of estimation errors $\left\{\mathbf{\epsilon}^{(i)}(k-N\,|\,k-N):i=1,2,...N\right\}$ while performing the computations over the interval $[k\text{-}N,\ k\text{-}N+1]$. Thus, while estimating $\mathbf{P}(k-N+1)$, new particles at the beginning of the interval $[k\text{-}N,\ k\text{-}N+1]$ are generated as follows

$$\hat{\mathbf{x}}^{(i)}(k-N\,|\,k-N) = \widetilde{\mathbf{x}}(k-N) + \mathbf{\epsilon}^{(i)}(k-N\,|\,k-N) \quad for \quad i=1,2,...N \quad (18)$$

where $\widetilde{\mathbf{x}}(k-N)$ represents the estimate generated by MHE problem solved over interval $[k\text{-}N,\ k]$. This modification facilitates *alignment* of the particles with the estimates generated using MHE.

4. Simulation Example

The performance of the C-EnKF based MHE algorithm is illustrated and compared with the MHE using a simulation example taken from the literature (Qu and Hahn, 2008). This example consists of a non-isothermal continuously stirred tank reactor model which includes reactor concentration (C_A), reactor temperature (T), and cooling jacket temperature (T_J) dynamics. The cooling water temperature is considered as the only measured variable. The horizon length used in MHE has 3 sampling times, with 20 samples used at each time point for the C-EnKF. State and measurement noise were considered zero mean Gaussian. The model parameters, initial values, and noise covariance matrices were taken from (Qu and Hahn, 2008)., The MHE problem was formulated using the simultaneous dynamic optimization formulation where the dynamic model is approximated with orthogonal collocation, and the resulting NLP problems were solved using IPOPT (see Zavala et al., 2008).

The evolution of the estimation errors is shown in **Erro! Fonte de referência não encontrada.**(a) where it is shown that the estimation errors for the C-EnKF-MHE are almost always smaller that the MHE. Figure 1(b) shows the sum of squared errors (SSE) for each state as a function of horizon lengths (from 2 to 6 sampling times) for both methods. Here, the SSE values have been normalized by the maximum SSE, which corresponds to the MHE using a horizon length of 2. From this figure it is easy to see that the performance of C-EnKF-MHE is significantly better than MHE for all the choices of horizon lengths. Moreover, there is hardly any change in the performance of C-EnKF-MHE when the horizon length is increased from 2 to 6. This improvement may be attributed to the fact that the approximation of the state error covariance matrix in the C-EnKF-MHE is statistically more consistent with bounds on states. Furthermore, at each sampling time the samples for the states used in the C-EnKF are not being redrawn. Instead, because the state samples are propagated over time, no assumption is needed on the distribution for the state estimation error. Moreover, unlike the

conventional MHE, the covariance estimation step does not involve linearization of the model as the covariance is approximated by Monte Carlo simulations. Additional computations arising from C-EnKF can be an issue for online implementation. However, when using the C-EnKF to approximate of the arrival cost, it is possible to achieve good estimator performance using even smaller horizon lengths. Working with shorter horizon lengths significantly reduces the on-line iterative computations. Also, since the measurements at the beginning of the horizon are known, the C-EnKF calculations can be done in the background between any two sampling times, thereby reducing the computation load at any sampling time.

(a)   (b)
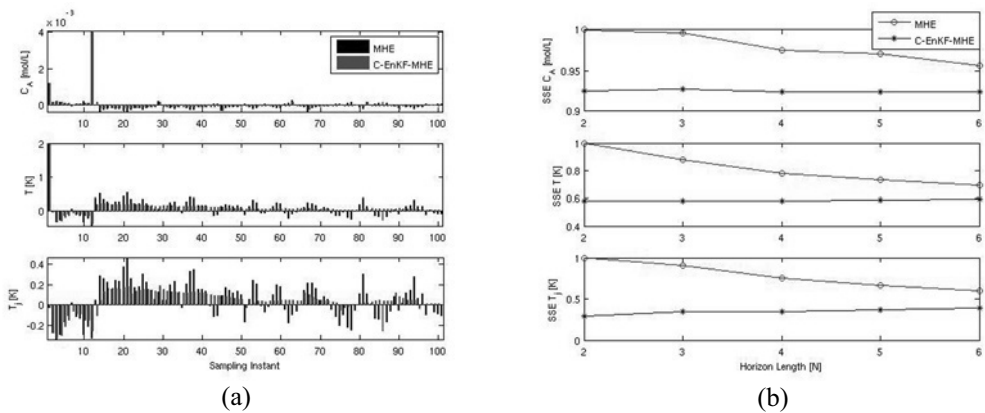
Figure 1: (a) State estimation errors. (b) Normalized SSE with as a function of horizon length.

5. Conclusions

A C-EnKF based scheme for updating the arrival cost in MHE is presented in this paper. It generates an approximation of the state error covariance matrix that is consistent with bounds on the state variables and without making assumptions on the distribution of the state estimation error. The simulation results demonstrate that the proposed formulation performs significantly better than the conventional MHE formulation. Moreover, the simulations reveals that MHE with short horizon lengths can be made to work well if the arrival cost estimation is improved.

References

V. M. Zavala, C. D. Laird, and L. T. Biegler. A Moving Horizon Estimation Algorithm Based On Nonlinear Programming Sensitivity. *Journal of Process Control*, In Press, 2008.

J. B. Rawlings and B. R. Bakshi (2006), Particle Filtering and Moving Horizon Estimation. *Computers and Chemical Engineering*, **30**, 1529–1541.

G. Evensen (2003) The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, *Ocean Dynamics*, 53, 343–367.

J. Prakash, S. C. Patwardhan and S. L. Shah (2008a), Constrained State Estimation Using the Ensemble Kalman Filter, *Proc. of the 2008 American Control Conference,* Seattle, June 2008.

C. C. Qu and J. Hahn. Computation of Arrival Cost for Moving Horizon Estimation via Unscented Kalman Filtering. *Journal of Process Control*, In Press, 2008.

Slug Control Structures for Mitigation of Disturbances to Offshore Units

Diego Di D. Pinto,[a] Ofélia Q. F. Araújo,[a] José Luiz de Medeiros,[a] Giovani C. Nunes[b]

[a]*Escola de Química, Universidade Federal do Rio de Janeiro. Av. Horácio Macedo, 2030, Ilha do Fundão, Rio de Janeiro, RJ, Brazil, CEP 21941-909*
[b]*PETROBRAS, ENGP/TPP/PMF, Av. Chile, 65 – Centro, Rio de Janeiro - RJ, Brazil, CEP 20035-900*

Abstract

Petroleum has been used by society in many ways, but mainly as an energy source. To meet the increasing energy demand, oil companies seek for new technologies to optimize production in a field. For this purpose, gas is employed for artificial lift, therefore enhancing productivity of old fields which suffer from insufficient pressure to force the oil out to the surface. Often, gas lift is associated to intermittent flow regime, slug flow that promotes intense disturbances to downstream offshore processes. In the present work, three control strategies for a gas-liquid separator are proposed: (i) feedback averaging level control; (ii) feedback + feedforward level control employing an ARX estimator; and (iii) override control on the production choke valve. The control schemes attempt to minimize downstream slug effects by exploiting the separator's hold-up, while imposing limits to liquid level and maintaining pressure under satisfactory performance.

Keywords: gas-lift, phase separators, slug flow, process control, ARX.

1. Introduction

Most nations rely on petroleum to meet their energy demand, requiring increased efficiency in the extraction process. In offshore oil fields, risers are used to transport a multiphase mixture (composed by oil, gas, water and sand) from the wellheads to separation tanks on producing platforms. In case of inexistence of separation and pumping facilities near the wellheads, on the sea bottom, this multiphase mixture must be propelled to the sea surface at expenses of the reservoir pressure. For deep water, a common situation is that the reservoir is not sufficiently pressurized to promote the flow of oil at satisfactory rates, requiring artificial ascension of oil.

The method is based on reducing the production column hydrostatic pressure by reducing its average density. The reduction of de density is given by injecting an elevating fluid, usually natural gas. Reducing this hydrostatic pressure will allow the reservoir pressure to force the oil out to the surface. The injection of the gas must be controlled; otherwise, it can promote pressure surges in the ducts and related equipments (Robinson *et al*, 2001). In some cases, the continuous gas lift, associated to high injection rate, is not economically efficient, requiring reduced injection rates responsible for intermittent flow regime (Santos *et al*, 2001). Figure 1 shows a gas lift process where: (i) gas is injected in the annular region; (ii) gas passes to the production

tube through a check valve located slightly before the reservoir; (iii) in the production tube, the gas is mixed with the oil reducing its density; (iv) as the density is reduced, the base hydrostatic pressure decreases; (v) being this pressure lower than the reservoir pressure, the column fluid ascension occurs pushed by the reservoir pressure; (vi) since the gas injection rate is low, the column ascension will carry the gas out from the column, increasing the density and, hence, the bottomhole pressure what will close the check valve, stopping the gas inflow; (vii) production ceases until enough gas is injected into the annular column, restarting the cycle.
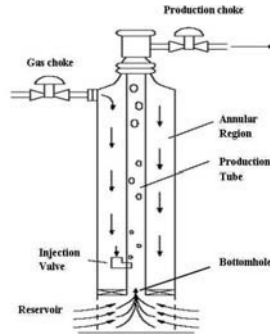
Figure 1: Gas Lift Process.

In consequence, high flow rates of gas have to be accommodated in the system of risers for further recovery by gas-liquid separators at the platform, before re-compression and re-injection. In this scenario, and depending on the flow conditions, a flow regime called *riser slugging* can develop in the pipelines. This regime, naturally highly transient, introduces severe oscillations of pressure and flow rates into the system. On the other hand, space and load constraints for offshore units favor compact equipments which increase sensitivity of downstream processes, mostly in the early stage of phase separation. Multiphase separators are, hence, a key-step in offshore facilities, with a two-fold process objective: (a) phase separation; and (b) dampening *riser slugging*, as in surge tanks. Faanes and Skogestad (2003) defined a *buffer tank* as a unit at which the holdup is explored to promote smooth operation. While in some level applications it is desirable to maintain tight regulation, such as in a reactor to achieving maximum capacity, this is seldom the case for level control in surge tanks, as it betrays the anti-surge requirement.

In the present work, three control strategies for a gas-liquid separator are proposed: (i) feedback averaging level control; (ii) feedback + feedforward level control employing an ARX estimator; and (iii) override control on the production choke valve. The control schemes attempt to minimize downstream slug effects by exploiting the separator's hold-up, while imposing limits to liquid level and maintaining pressure under satisfactory performance.

2. Process Model

For testing the control strategies and developing the stochastic predictor, a process model is developed based on Aamo *et al*, 2005. The dynamic behavior of the gas lift derives from the interaction between the annular and tubular volumes above the gas injection point, described by mass balances shown in Eqs. 1, 2 and 3, where *H* and *W,*

are, respectively, hold-up (kg), mass flow rate (kg/s). Subscripts *A, G, L, I, C, R* and *P,* stand respectively for annular region, gas in the production tube, liquid in the production tube, injected gas in the production tube, gas in the production choke, reservoir, production choke.

$$\frac{dH_A}{dt} = W_G - W_I \tag{1}$$

$$\frac{dH_G}{dt} = W_I - W_C \tag{2}$$

$$\frac{dH_L}{dt} = W_R - W_P \tag{3}$$

The valve responsible for injecting gas into the annular region is taken as a *swing-check-valve*. The choke valve for feeding gas to the annular region and the production choke valve are modeled as gate valves, assuming that the oil reservoir and the offshore separator are at constant pressure. The oil single phase flow section between the reservoir and the gas injection point is solved in pseudo-stationary mode with the reservoir and injection point pressures, assuming incompressible flow with friction factor given by Chilton's Equations, which comprise all hydraulic regimes.

The gas phase behavior is simplified to an isothermal ideal gas. Furthermore, temperatures in the gas lift subsystems are simplified to annular temperature equal to fed gas temperature, and production tube temperature identical to the oil temperature in the reservoir.

Figure 2a shows gas flow profiles for gas lift simulated at conditions shown on Table 1. Fixing these conditions and varying the gas injection rate leads to the identification of two distinct operating regions: (i) continuous gas lift or stable region; and (ii) intermittent gas lift or unstable region, as displayed on Figure 2b.

Table 1: Gas Lift Conditions.

	Oil	Gas
Pressure, Temperature	160 bar, 108°C	120 bar, 60°C
Density , Viscosity	850 kg/m³, 0.012 kg/m.s	Ideal Gas
Choke Valve (opening fraction)	0.8	0.5
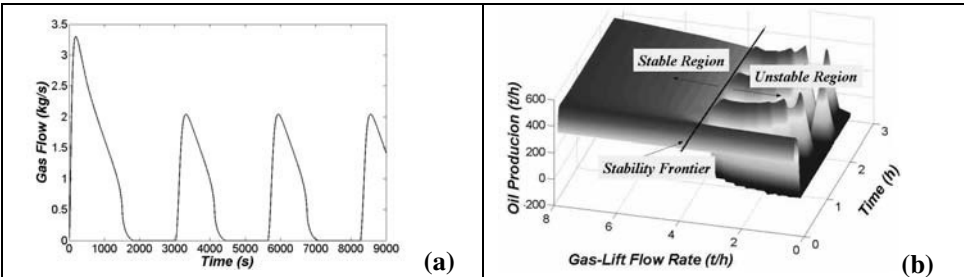
Figure 2: Simulation Results. (a) Gas Flow Profile for Gas Lift Simulated at Table 1 Conditions, (b) Sensitivity Analysis to Injection Gas Flow Rate

The gas lift model was expanded to include the operation of the gas-liquid separation. Mass balance for the liquid and gas phases yield Equations 4 and 5, where *L, G, C, D*

and *V* are respectively the liquid flow rate, the gas flow rate, separator's length, separator's diameter, and the volume. Subscripts *I, o, T* and *L* stand for, respectively, input rate, output rate, total, and liquid.

$$\frac{dh_L}{dt} = \frac{L_i - L_0}{2C\sqrt{h_L(D-h_L)}} \tag{4}$$

$$\frac{dP}{dt} = \frac{P(L_i + G_i - L_0 - G_0)}{V_T - V_L}; \; V_T = C\frac{\pi D^2}{4} \tag{5}$$

3. Control Strategies

The control objectives are to maintain separator's pressure and its level in an appropriate range. Since pressure in the separator affects flow rate through the valves, it will impact the dynamics of the gas lift, and decrease the oil production. The level controller protects the separator against overflow or drainage. Also, due to the eventual occurrence of severe slug caused by the intermittent gas lift, level controller must provide buffer capacity, protecting downstream operation. For this purpose, three strategies are tested: (i) averaging level control exploiting the liquid hold up of the tank; (ii) feedback and feedforward control using an ARX predictor of separator feed flow rate; (iii) override control on the production choke valve and feedforward control. All the three strategies are shown schematically in Figure 3.
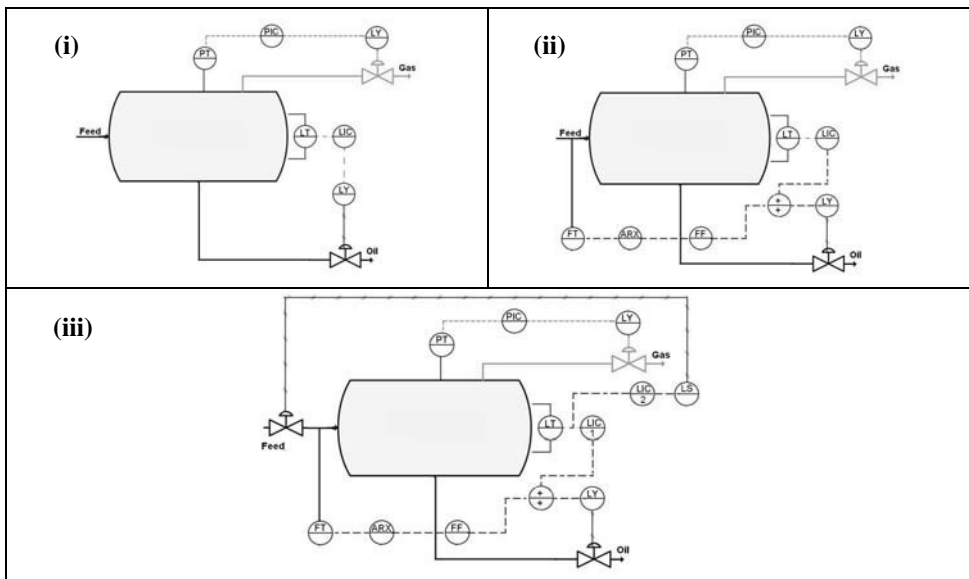
Figure 3: Control Strategies for Gas-Liquid Separators.

3.1. Averaging Level Control

Cheung and Luyben (1980) explore different level control strategies, such as the strategy proposed by Shunta and Fhervari (1976), the wide-range controller, with adaptation of PI controller tuning parameters (K_C and τ_I) according to Equation 6:

$$K_C = f \ K_{C0}; \quad \tau_I = \frac{\tau_{I0}}{f} \tag{6}$$

In this work, an alternative adaptation law proposed *f*, by Araújo *et al* (2007), is adopted, as given in Equation 7.

$$f = 1 + \frac{1}{1 + e^{(\lambda_1(E_1 - abs(e)))}} + \alpha \left(1 + \frac{1}{1 + e^{(\lambda_2(E_2 - abs(e)))}}\right) + \beta \left(1 + \frac{1}{1 + e^{(\lambda_3(E_3 - abs(e)))}}\right) \tag{7}$$

3.2. Feedforward + Feedback Control

For this strategy, an identification of the gas lift process is required as separators inflow is frequently not available from measurements. An ARX 3x3 with 3 inputs (gas injection rate, choke opening fraction, and separator tank pressure) and 2 outputs (oil and gas flows) is presented in Pinto (2009) and is used to inferring oil and gas flows. Faanes and Skogestad (2003) propose a model for buffer tank level response (*h(s)*) to inflow rate (*d(s)*, Equation 8) that is herein applied to a separator to obtain the feedforward control law (Equation 9). with $K = -16$ and $\tau = 343.53$ s. The feedback controller is identical with same as showed at strategy 3.1.

$$h(s) = \frac{1}{\tau s + 1} d(s) \tag{8}$$

$$G_{FF}(s) = \frac{K}{\tau s + 1} \tag{9}$$

3.3. Override Control

In this strategy, under normal operating conditions, level controller (LIC1) is adapted according to Equation 7. Under severe slug, this controller is overridden though a low value switcher (LS) to a level controller (LIC2, with set-point equals to a high level limit) which acts reducing opening of the production choke valve.

Figure 4 compares the three strategies.
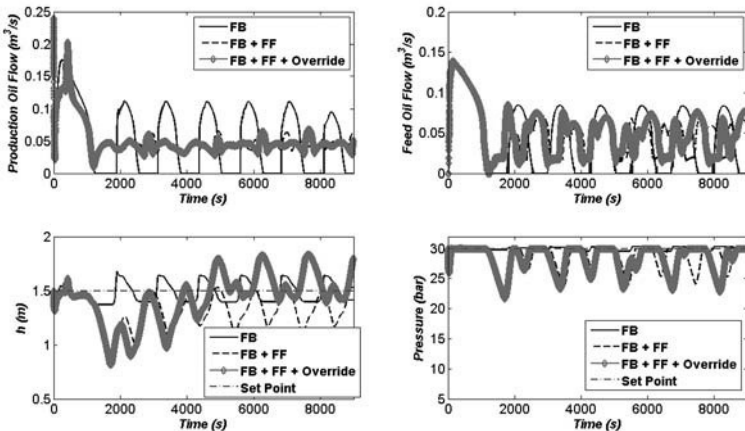
Figure 4: Comparison of Control Strategies.

Feedback level controller with adaptation law given by Equation 7 is able cope with slugs but is unable to reduce flow perturbations propagating downstream. With a feedforward combined to the adaptive feedback, it is possible to minimize the slug effects, while maintaining level and pressure in the separator within an acceptable working range. The override control is used as a security layer, preventing the separator's level from reaching dangerous values. It's possible to observe in Figure 4 that the results from both feedforward and override control (strategies ii and iii) are almost the same differing only on the level control, where the addition of the override strategy shows improves performance.

4. Conclusions

To mitigate the impact of slug flow downstream of a gas-liquid separator, this work approaches 3 control strategies. These strategies are presented to work in layers: a feedback layer, for correcting adaptively level deviations, a feedforward layer to anticipate control action based on an ARX predictor of the disturbance, the inlet liquid flow), and adopting a new level controller through override of the feedback controller.
It's possible to conclude that strategies (ii) and (iii) improve performance of an adaptive feedback controller. It is worth noting that the override layer only interferes when the separator's level reaches dangerous values.

References

AAMO, O. M., EIKREM, G.O., SIAHAAN, H.B., FOSS, B.A., 2005, Observer design for multiphase flow in vertical pipes with gas-lift – theory and experiments, Journal of Process Control, 15, 247-257.

ARAÚJO, O.Q.F., MEDEIROS.J. L., NUNES, G.C., 2007, Modeling and Control Solutions for Riser Slugging in Offshore Oil Fields, Proceeding of European Congress of Chemical Engineering.

CHUNG, T.F., LUYBEN, W.L. ,1980, Nonlinear and Nonconventional Liquid Level Controllers, *Ind. Eng. Chem. Fundam.*, 19, 93-98.

FAANES, A., SKOGESTAD, S., 2003. Buffer Tank Design for Acceptable Control Performance. *End. Eng. Chem. Res.*, v. 42, 2198-2208.

PINTO, D. D., 2009, Estratégias de Controle Contra Intermitência Severa na Alimentação de Separadores Offshore, MSc. Thesis, Escola de Química, Federal University of Rio de Janeiro.

ROBISON, C. E. , SCHNATZMEYER, M. A., Bayh III, R. I. , 2001, Self-regulating lift fluid injection tool and method for use of same. U.S. Patent No. 336459.

SANTOS, O. G., BORDALO, S. N. , ALHANATI, F. J.S.. 2001, Study of the dynamics, optimization and selection of intermittent gas-lift methods—a comprehensive model, Journal of Petroleum Science and Engineering, 32, 231 – 248.

Embedded Control and Monitoring Systems in Production Machine Networks

Sirkka-Liisa Jämsä-Jounela,[a] Mikko Huovinen [b]

[a]HelsinkiUniversity of Technology, Department of Biotechnology and Chemical Technology, P.O.Box 6100, FI-02015 Espoo, Finland
[b]Tampere University of Technology, Department of Automation Science and Technology, P.O.Box 692, FI-33101 Tampere, Finland

Abstract

The ongoing globalization trend is tightening competition and setting new, higher efficiency requirements in the process industries. In order to enhance the efficiency of the production chain, new functionalities and information networking must be incorporated into the production equipment. This paper proposes a concept of equipment automation that utilizes new information and communication technologies and more flexibly adds equipment intelligence. The feasibility and applicability of the concept is demonstrated via case studies of a grinding circuit and a pulp drying process.

Keywords: Networked equipment automation, intelligent process equipment.

1. Introduction

The globalization process is tightening competition and setting, higher efficiency requirements in the process industries. In addition, factories now have to be considered as a part of sustainable development. Energy savings and pollution prevention have become priorities. Safety and maintenance are also listed as the main equipment design parameters. At the present time, however, the intelligence level of the process equipment and their networks is still low. In order to enhance the efficiency of the production chain, new functionalities and information networking must be incorporated into the production equipment. This can be further enhanced by applying embedded automation middleware to manage complex networked machines and distributed resources.

The goal of automation middleware is to increase the intelligence level of the production machines by networking and integrating the embedded systems of the machines into larger automation system entities. The functionalities embedded in the middleware cover the management of the whole process system at the life-cycle scale. The methods can utilize connections to higher-level systems and remote resources, as well as to the information-refining tools provided by the middleware platform. The middleware thus integrates lower level process resources, and further refines the collected information into a more convenient form that is subsequently presented to the users and/or to other information systems, such as MES or ERP.

The purpose of this paper is to propose a common way of incorporating the current communication technologies, and especially the communication protocol and network topology design, into process equipment in such a way that it facilitates the implementation of new methods and subsequently satisfies the efficiency requirements.

2. Survey of Industrial Automation Architecture and Communication Technologies

The evolution of communication technologies has had a strong influence on changes in the structure of industrial automation systems. Up until now, communication support in plant automation systems has been defined according to the Computer Integrated Manufacturing (CIM) concept. In this hierarchical structure, levels of functionality are identified in such a way that each device is designed for a specific task, and specific networks are used to interconnect devices at the same level, i.e. running the same task.

However, the devices have recently started to include more than one function, or module, which increases the intelligence level of the equipment automation. Devices like sensors that have traditionally been used for measurement now have to support e.g. maintenance or monitoring tasks. This means that the traditional hierarchical structure is increasingly being replaced by a distributed communication architecture. Nevertheless, the hierarchical structure still exists - and this is also advisable - in most of the process control strategies and plant automation systems, as illustrated in Fig. 1 and 2. This phase in the evolution of automation systems has been called FCS (Field Control System).
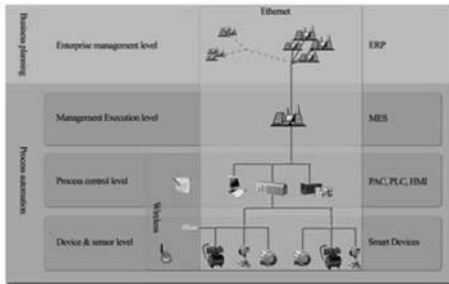
Fig. 1. Hierarchy of the plant automation.   Fig. 2. Future scenario of the plant automation
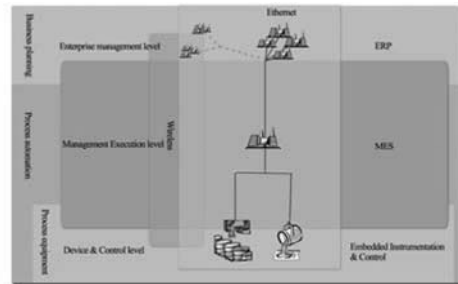
2.1 Low-layer communication protocols
Currently the most widely available industrial networks can be classified into three main categories: traditional fieldbusses, Ethernet-based networks and wireless networks. A comparison of the key properties of the currently, most widely available networks in each of the three main categories is given in Table 1.

2.2 High-layer data specifications
The information standards for process operation and maintenance are driven by OpenO&M Initiative joint working groups, which mainly represent three industrial organizations: MIMOSA, the OPC Foundation, and ISA's SP95. One of the most strongly established standards, the OPC, also enables the use of state-of-the-art technologies such as web services. On the other hand, the traditional fieldbusses (like Profibus or Foundation Fieldbus) have defined concepts for the manufacturer-independent integration of field devices such as FDT/DTM.

3. Proposed Architecture for the Networked Equipment Automation

3.1 Components of the automation concept
The concept of equipment automation must be applicable over a wide range of different industrial processes. The main starting point is to bring the automation, with intelligent software-based functionalities, near to the process equipment while, on the other hand,

designing a communication network that robustly provides the services that are already incorporated. Therefore, the automation devices for one specific piece of equipment or sub-process are networked into one process system node, thereby providing connectivity to other parts of the process. The equipment 'intelligence' is embedded as a part of the automation devices, using software-based modules (Fig.3.)
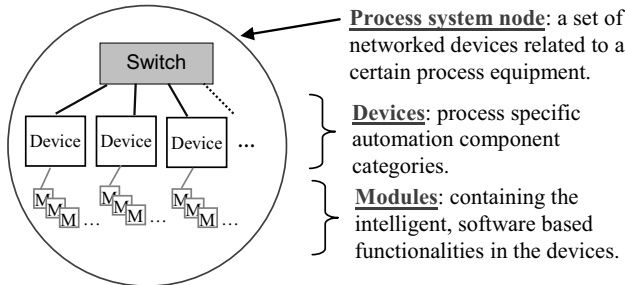
Fig. 3. Structure of the process system node.

The process system nodes are further connected to the supervisory automation level with a high bandwidth network (Fig. 4). In addition, the nodes can also provide a wireless connection link. Finally, in order to implement the equipment's advanced operating algorithms and procedures, a data processing unit (PLC or soft-PLC) is needed as a part of the process system node. This leads to a scheme in which the equipment automation provides the intelligent operating functionalities related to a specific task, which typically means one unit operation in the process chain. Compared to the traditional DCS hierarchy, the addition of intelligence offers a more dynamic platform for autonomous systems. Intelligent devices can act more independently and the improved communications enable interaction between the distributed assets. In fact, the FCS hierarchy provides a new platform for the implementation of agent technologies in the process industry.

3.1. Communication network topology

The architecture of the concept relies on a *backbone* that is capable of providing the main connectivity between the different devices. Moreover, it includes the connection point to remote resources via the Internet. The process system nodes, which are linked to the backbone, form sub-networks. These provide supplementary networks for establishing the instrumentation power supply and wireless links. It also manages the mutual interconnections between local devices. Redundancy is also provided for the backbone. The network system topology is presented in Fig. 5.

The communication protocol applied in the backbone should be able to link with office and remote resources and, as a result, have the same low layer technology as used in Internet networks, to provide a large bandwidth to integrate all the sub-networks, and also to support several high layer technologies, such as web and multimedia applications. The use of *Ethernet* networks is therefore proposed. The Ethernet technology is applied in the process system node network level. The connection to the devices is shared by using *switch* in order to improve the time determinism and isolate the local traffic. The Ethernet also has the advantage of facile connectivity to traditional fieldbusses. Additionally, Ethernet ensures device interoperability by supporting open standards, such as OPC.
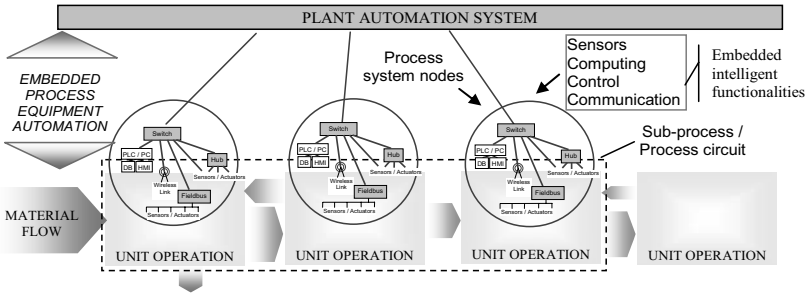
Fig. 4. Concept of the equipment automation level between the physical process and the supervisory plant automation system.
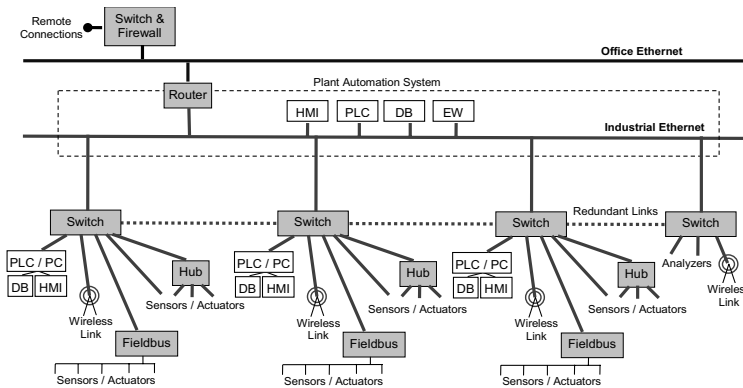
Fig. 5. Topology of the equipment automation network architecture.

Table 1. Comparison of the currently, most widely available industrial networks

Type	Protocol	Manufacturer	Throughput	Cycle Time	Range (Length)	Number of Devices	Energy Efficiency
Traditional Fieldbus	CAN	Bosch	10kb/s - 1Mb/s		40m (1Mb/s), 5km (10kb/s)	max. 32	
	WorldFIP(1158-2)	Schneider	31,25kb/s, 1Mb/s, 2,5Mb/s (5Mb/s optical fiber)	2 ms, 5 ms	1km (1Mb/s), max 4km	max. 256, 64/seg	Power over network solution
	Profibus DP	Siemens	9,6kb/s - 12Mb/s	depending on configuration <2ms	100m (12Mb/s), 1,2km (9,6kb/s)	max. 126	
	Profibus PA	Siemens	31,25kb/s	depending on configuration <2ms	1900m	max. 32/seg	Power over network solution
	DeviceNet			2,0 ms, 4,2 ms, 10 ms			
	ControlNet	Rockwell Automation	5Mb/s	<0.5 ms	5km (5Mb/s), 30km+ optical fiber	max. 99	
	Interbus	Phoenix Contact	500kb/s	1,8 ms, 7,4 ms, 140 ms	~13km	max. 512	
	AS-i	Siemens	167kb/s	4,7 ms	100m/seg, 300m	max. 64	Power over network solution
	Foundation Fieldbus H1	Fieldbus Foundation	31.25kb/s	36 ms,100 ms, 600 ms	max 1900m	max. 32	Power over network solution
Type	Protocol	Manufacturer	Throughput	Cycle Time	Range (Length)	Number of Devices	Energy Efficiency
Ethernet Based	Ethernet/IP	Rockwell Automation	10Mb/s, 100Mb/s, 1Gb/s	>10 ms			
	Foundation Fieldbus HSE	Fieldbus Foundation	100Mb/s		100m/seg		
	EtherCat	Beckhoff	100Mb/s	<50 µs	100m/seg	~65535	Power over network solution
	Powerlink	B&R	100Mb/s	<500 µs	100m/seg		
	ProfiNet IO	Siemens	100Mb/s	1 ms, 10ms	100m/seg		
	Sercon III	Bosch Rexroth	100Mb/s	31,5 µs	100m/seg	max. 254	
	Modbus TCP	Schneider	10Mb/s, 100Mb/s, 1Gb/s	>10 ms			
Type	Protocol	Manufacturer	Throughput	Cycle Time	Range (Length)	Number of Devices	Power Consumption
Wireless	IEEE 802.11 (b/g/n)		54Mb/s, 100Mb/s		30-100m		Medium
	Bluetooth, IEEE 802.15.1		1Mb/s		10 m	7	Low
	IEEE 802.15.4 /ZigBee		20kb/s, 40kb/s, 250kb/s		10 m	~65k	Very Low
	WirelessHart	Hart Communication Foundation	250 kb/s		10 m		

4. Add–On Service Modules of the Intelligent Process Equipment

Intelligent process equipment in modern automation systems can also be thought of as forming a functional hierarchy. The equipment provides *services* designed to meet the higher level *objectives*. There are production management services consisting of daily operations, abnormal event management, operator support and training, product data and lifecycle management services. All these consist of lower level *activities*, such as process control, performance monitoring and field management. The implementation of the activities consists of *tools and routines* as presented in Fig. 6.
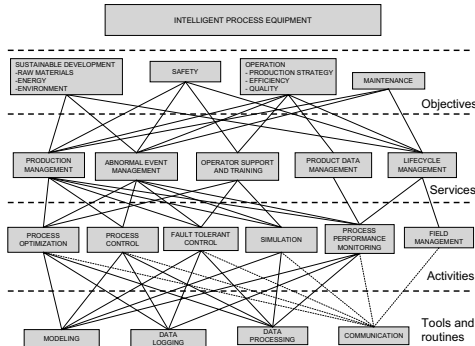
Fig. 6. Functional hierarchy of the intelligent process equipment

5. Case Examples

5.1. Case example: applying the concept of equipment automation to a grinding circuit
The proposed equipment automation concept is applied to an example of a mineral grinding process. It is proposed that the instrumentation of the process is grouped into four categories, typical for the process. These are: basic sensors, machine vision, process analyzers, and data processing units.

The process system nodes are divided according to the data processing units for the operation and monitoring of the mill, pump and hydrocyclone. The fourth node is reserved for an elemental analyzer, which is located physically further from the grinding circuit. One of the most important nodes, which includes the 'mill automation' data processing unit, has a wireless link enabling connection to the particle size and elemental assay analyzers. Other wireless sensors located in problematic points, such as the rotating mill shell, can also be added to the system. The industrial Ethernet connection is favored for all the instrumentation connections, as illustrated with the 'machine–vision' device in the figure. The network topology of the equipment automation in this case is illustrated in Fig. 7.

5.2. Case example: applying the concept to monitor a pulp drying process
The concept was also implemented and deployed in the drying section of a Finnish pulp mill. The application is primarily intended to be used for production management services through process optimization and performance monitoring tools. It is also useful for abnormal event management, lifecycle management and operator support. The tools available for process optimization and performance monitoring include clustering, conditional histogram and control loop monitoring. An extensive performance measurement system has also been developed. The measurement system is hierarchical representing the whole system, subprocesses, control loops and the

equipment. On the equipment level features of intelligent field devices were exploited while on the control loop level a control loop monitoring tool called LoopBrowser was used. Examples of subprocess measurements are relative energy consumption and relative variations of controlled variables. The whole drying process was represented by an OEE indicator.
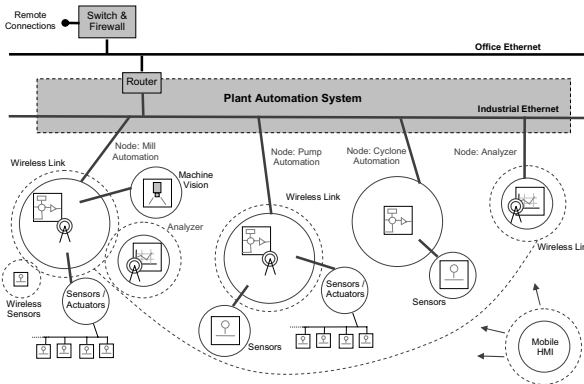
Fig. 7. Simplified abstraction of the network topology, with device connections for the process.

The clustering method is based on a modified K-means algorithm for identifying operating points and for detecting abnormal situations. The algorithm was chosen based on its computational simplicity enabling easy online implementation. The information is also used for diagnostic purposes because the application provides a contribution plot for showing the deviations of the input variables from their typical values. The conditional histogram method is used for constructing a database representing normal performance in each operating point. The method also includes routines for detecting whether the current measured performance deviates from the normal performance. The control loop monitoring system produces data about the performance of the monitored control loops. Four performance indices are used, which basically measure the accuracy and stability of the control loops.

The refined information serves the lifecycle module by providing monitoring performance throughout the equipment lifecycle. Intelligent field devices can be incorporated to deepen the scope of the application through embedded device level monitoring, as well as diagnostics which provide valuable information about the performance and ambient conditions of the field devices. For example, valve positioners typically provide control error and stiction measurements that can be used for determining whether a detected abnormal situation originates from the monitored valve-actuator-positioner package or from an outside source. Efficient exploitation of field device level information requires the use of device integration technologies, such as FDT/DTM. The overall performance measurement system was designed to give a comprehensive picture of the overall state of the whole process through the use of sub-process performance indicators. These measure e.g. relative variations in the controlled variable and efficiency measures.

6. Conclusions

In order to meet the tightening performance requirements in the process industries, the efficiency of the process equipment must be improved. Therefore, an equipment automation concept was formulated and proposed The concept outlines the procedure for efficient utilization of the latest information and communication technologies.

1317

Multivariate Statistical Control of Emulsion and Nanoparticle Slurry Processes Based on Process Tomography, Dynamic Light Scattering, and Acoustic Sensor Data

Rui F. Li,[a] Lande Liu,[a] Xue Z. Wang,[a] Richard Tweedie,[b] Ken Primrose,[b] Jason Corbett,[c] Fraser McNeil-Watson[c]

[a] Institute of Particle Science and Engineering, University of Leeds, Leeds LS2 9JT,UK
[b] Industrial Tomography Systems Ltd, Speakers House, 39 Deansgate, Manchester M3 2BA, UK
[c] Malvern Instruments Ltd, Enigma Business Park, Malvern, Worcestershire WR14 1XZ,UK

Abstract

This paper describes the use of multiple on-line sensors including electrical resistance tomography (ERT), dynamic light scattering (DLS) and ultrasound spectroscopy (USS) for real-time characterization of process operations processing emulsions and nanoparticle slurries. The focus is on making novel use of the spectroscopic data to develop multivariate statistical process control (MSPC) strategies. The ERT data at different normal operating conditions was processed using principal component analysis and used to derive two MSPC statistics, T^2 and SPE (squared prediction error) for detecting abnormal changes in mixing conditions. The corresponding particle size distribution was monitored using DLS and USS. Two case studies, a sunflower oil–water emulsion system and a silica suspension system, were examined.

Keywords: dynamic light scattering, acoustic spectroscopy, process tomography, multivariate statistical process control.

1. Introduction

An experimental system for studying on-line sensing techniques for nanoparticle slurries and emulsions has been built at the University of Leeds. The rig is equipped with dynamic light scattering (DLS) and ultrasound spectroscopy (USS) instruments for particle size and z-potential measurement, as well as electrical resistance tomography (ERT) for characterising the mixing conditions and solid concentration. In addition, flowrate, pH and temperature are also measured. This paper describes a novel method for making effective use of the spectroscopic to develop multivariate statistical process control (MSPC) charts.

2. Materials, Instruments and Experiments

2.1. Materials
Two materials were investigated in this study, a sunflower oil – water emulsion and a silica suspension. The emulsion was manufactured by a membrane emulsification rig in our lab. The oil volume concentration is 33%, and the mean size of droplets is around

700 nm. The silica suspension, Nissan SnowTex ZL, of 24.0% vol solid concentration and particle mean size between 70 nm – 100 nm was   purchased from Nissan America Incorporation.

2.2. Experimental Rig and Instruments

Figure 1 shows the flow diagram of the experimental rig. It consists of mainly two parts, a 10 litre jacket reactor and a circulation loop. The 10 litre reactor is used as a buffer tank to store the multi-phase fluid. It is equipped with sensors of temperature, pH and conductivity. A Julabo water bath (FP50) is used to control the temperature inside the tank. Mixing in the tank was achieved using a pitched-blade stirrer. A mono pump with flow control is used to feed fluid to the circulation loop. A flow metre and a glass window were also fitted into the loop for visual observation of the flow. The ERT and USS systems were directly connected to the circulation pipe. A by-pass pipe with flow control and indication was designed for connection of the DLS sensor.
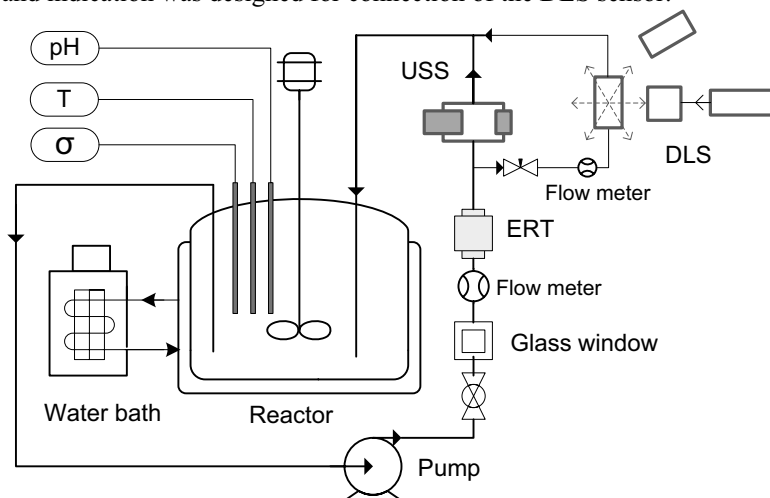
Figure 1. The experimental rig

The ERT instrument was provided by Industrial Tomography Systems Ltd [1]. It has 16 electrodes and can make 104 voltage measurements in approximately 25 ms. The tomography images are reconstructed from the 104 voltage measurements using a reconstruction algorithm. Both USS and DLS are supplied by Malvern Instruments Ltd [2].

2.3. Experiments

The purpose of the experiments is to investigate the mixing performance and measure particle size for two particulate processes at various concentrations and flowrates. During the experiments, in order to obtain different concentrations, the original emulsion, or suspension was gradually diluted to lower concentrations by adding sodium dodecyl sulfate surfactant solution or water into the concentrated emulsion or suspension. Under each concentration of the emulsion or suspension, different flowrates to the circulation loop were used to study the effect of flowrate on the mixing performance and particle size. During the experiments, tomography measurement was recorded at a rate of 60 frames per minute. At each concentration and each flowrate, one acoustic measurement was recorded by USS. In addition three size and three zeta-potential measurements were taken by the on-line DLS at every concentration and every

flowrate. Table 1 summarises the concentrations, flowrates and the numbers of each type of measurement obtained during the experiments for the two materials.

Table 1. Summary of measurements

Material	Volume concentration	Flowrate (l/s)	Number of DLS particle size measurement	Number of DLS zeta potential measurement	Number of USS particle size measurement
SunflowerOil – Water Emulsion	33%, 30%, 25%, 20%, 15%, 10%, 5%, 3%, 2%, 1%	0.190, 0.166, 0.133, 0.108, 0.083, 0.055, 0.041	3×10×7	3×10×7	10×7
SnowTex-ZL Suspensions	24%, 20%, 16%, 12%, 8%, 4%, 2.3%, 1.6%, 0.8%.	0.190, 0.166, 0.133, 0.108, 0.083, 0.055, 0.041	3×9×7	3×9×7	9×7

3. Multivariate Statistical Process Control Based on ERT Measurements

A significant development in multivariate statistical process control (MSPC) in recent years has been due to the introduction of principal component analysis (PCA) and Independent component analysis (ICA) for compression of data prior to deriving the MSPC's two statistics[4,5], Hotelling's T^2 and SPE. If PCA is used, the latent variables, i.e., the first few PCs are used rather than the original variables. The Hotelling's T^2 can then be calculated by

$$T^2 = \sum_{i=1}^{k} \frac{t_i^2}{S_{ti}^2} \tag{1}$$

where S_{ti}^2 is the estimated variance of t_i, k the number of PCs used, and t_i the ith score. The second statistic SPE (squared prediction error) model is:

$$SPE = \sum_{i=1}^{m} (y_{new,i} - \hat{y}_{new,i})^2 \tag{2}$$

where $\hat{y}_{new,i}$ and $y_{new,i}$ are the reconstructed value of the ith process variable from the reference PCA model and the measured value of this variable respectively. SPE is also refereed to as Q statistic or distance to the model. It represents the squared perpendicular distance of a new multivariate observation from the projection space. The procedure for developing the tomography based MSPC is described below.

The first step is the collection of normal operational data. The data is then pre-processed using two scaling methods. Since the tomography measurements (in this case the measurements are a set of voltages between different electrodes) are dependent on the conductivity and concentrations of fluids in the pipeline, it is necessary to remove the effect of these parameters on the measurements so that the developed MSPC model can be generalised and is independent of the concentrations and conductivity of the fluids.The first scaling method, called scaling over a frame, is applied to each frame of the collected data:

$$V_{new}(i,j) = \frac{V_{raw}(i,j) - V_{min}(i)}{V_{max}(i) - V_{min}(i)} \quad (j=1,2,3,\ldots,104) \tag{3}$$

where $V_{new}(i, j)$ and $V_{raw}(i, j)$ are the jth pre-processed and raw voltage measurement on the ith frame in the collected data, $V_{min}(i)$ and $V_{max}(i)$ the minimum and maximum raw measurement values on the ith frame, respectively. The second method for pre-processing the data is auto scaling to make each measurement have the same variance.

In the third step, the pre-processed data is analysed using PCA. A smaller number of latent variables, i.e. principal components (PCs) are derived. This transformation reflects the correlations between the tomography measurements at normal mixing states. The T[2] and SPE control limits at different confidence levels are derived.

Finally, the MSPC model is applied to new tomography data for mixing condition monitoring. It needs to point out that all new measurements were also pre-processed using the two scaling techniques described above. If the SPE is over the control limits, it could be an indication that the current mixing state had gone out of or was very different from the predefined normal mixing states. If the SPE is in the control limit and T[2] out of the limit, this could be an indication of bad mixing. After poor mixing is detected, contribution plots can be applied to analyse the key measurements contributed to the abnormal mixing.

4. Results and Discussions

4.1. Sunflower-oil – water emulsion

The ERT voltage measurements of sunflower oil – water emulsion, at flowrates 0.166, 0.133 and 0.108 l/s of concentrations from 33% to 5%, were used to develop the PCA based MSPC model. There are 23768 frames in total. The voltage data were firstly pre-processed using the scaling methods described in the previous section. Then PCA was applied to the pre-processed data to develop the MSPC models. 20 PCs were selected which explained about 73% variance of the data. Based on the selected PCs, control limits of two statistics SPE and T[2] at different confidence levels were calculated. The MSPC models were applied to monitor the mixing performance at different operating conditions. Figure 2 shows an example of emulsion at 20% concentration and various flowrates. It can be seen that significant increases of both SPE and T[2] occurred after the flowrate was reduced to 0.055 and 0.041 l/s. This indicates that the mixing became poor at these flowrates. The reconstructed conductivity tomography during the flowrate change, shown in the Figure 3, confirmed the deteriorating mixing at reduced flowrates.

(a)   (b)
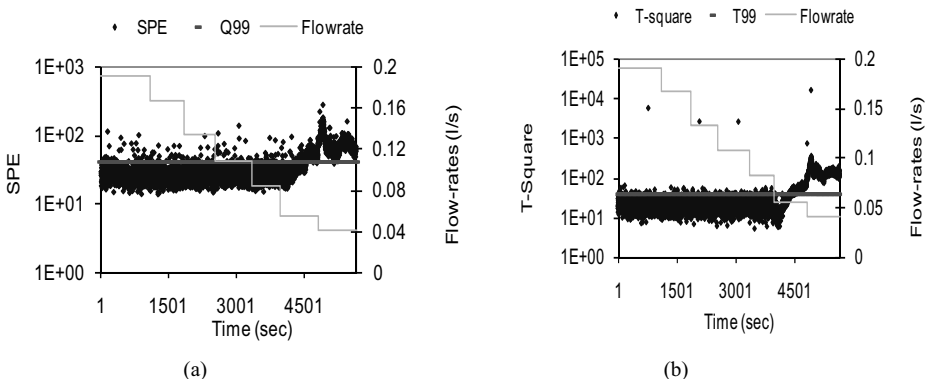
Figure 2. SPE and T[2] Monitoring charts for the sunflower oil – water emulsion system. The concentration is fixed at 20%, while the flowrate is decreased. Q99 – control limit is 99%.

| 3948 | 3955 | 4048 | 4148 | 4248 | 4348 | 4448 | 4548 |

| 4648 | 4748 | 4848 | 4948 | 5048 | 5148 | 5248 | 5348 |

2.1386 ms ▬▬▬▬▬▬▬▬▬▬▬▬▬▬ 2.2012 ms
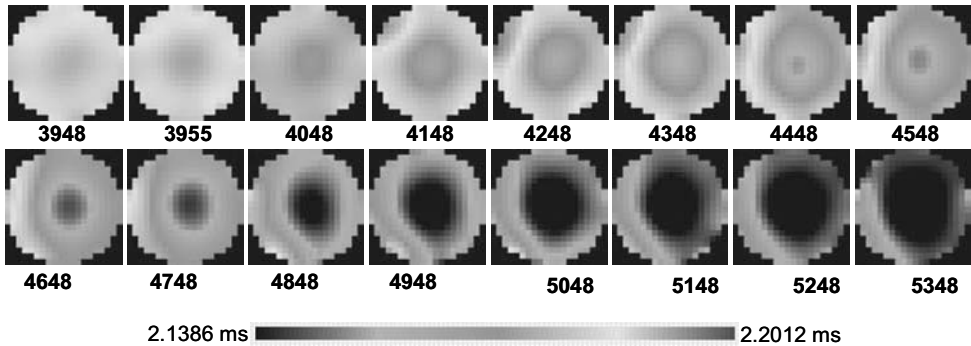
Figure 3. Reconstructed conductivity tomography images for the sunflower oil – water emulsion system. Concentration is fixed at 20%, but flowrate changes.
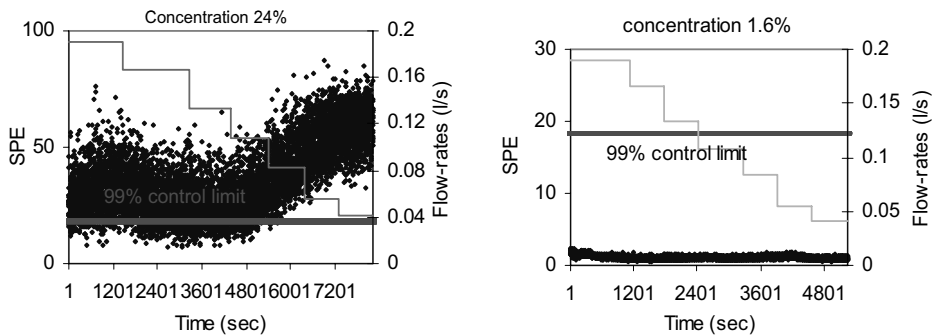
Figure 4. SPE charts for silica suspension at solid volume concentrations of 24% and 1.6%.

4.2. Mixing performance of SnowTex-ZL silica suspension

MSPC models were developed based on the tomography measurements at flowrates of 0.166, 0.133 and 0.108 l/s, for all concentrations studied. Twenty principal components were chosen, capturing about 88% of the variance.

The MSPC models were applied to monitor the suspension at different concentrations and flowrates. Figure 4 shows the SPE control charts at concentrations 24% and 1.6%, which indicates that the higher the concentration, the poorer the mixing.

4.3. Particle size measurements by on-line Zetasizer and Ultrasizer

Particle size distribution was measured using DLS and USS for both materials at different concentrations and flowrates. For the silica suspension, both DLS and USS gave consistent results when the flowrate was changed while concentration remained constant, as indicated by figures 5(a) and 5(b). However with the increase of solid concentration, the mean size measured by the DLS is increased, while the mean size measured by the USS is decreased, as shown in figure 5 (b) and (c), reflecting the effects of solid concentration on the measurement results.
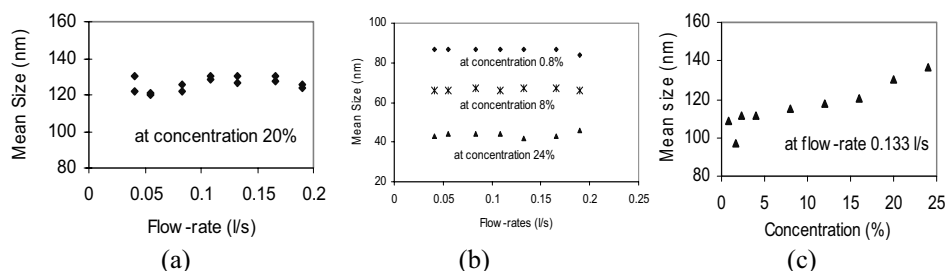
Figure 5. Particle mean size measurements. (a) DLS result at a fixed concentration; (b) USS result at different flowrates and concentrations, and (c) DLS result at a fixed flowrate but varied solid concentrations.

5. Final Remarks

Nanoparticle processing in industry faces major challenges in process scale-up and in maintaining consistency and reproducibility in product quality. Online characterization of product quality and process conditions can help industry effectively address the challenges, because based on real-time measurements, process control and quality assurance strategies can be developed. The work has investigated the use of dynamic light scattering (DLS), ultrasound spectroscopy (USS) and electrical resistance tomography (ERT) for online characterization of processes of nanoparticle suspensions and emulsions. The system provides real-time information of particle size distribution (PSD), zeta potential, conductivity and solid concentration and visualizes the mixing behaviour. In addition, ERT provides solid concentration data that is needed by DLS and USS for PSD characterization. This study proposed the use of multivariate statistical process control, or MSPC for making use of the data for process monitoring and control. Two cases studies proved the feasibility of the proposed approach. The combined sensors and the MSPC system together provide a platform for on-line characterization of emulsions and nanoparticle slurries.

6. Acknowledgement

References

1. Industrial Tomography Systems. http://www.itoms.com. Accessed February 2009
2. Malvern Instruments Ltd. http://www.malverns.com. Accessed Februatry 2009.
3. MacGregor J. F., T. Kourti, Statistical process control of multivariate processes. Control Engineering Practice, 1995, **3:** 403-414.
4. Albazzaz H., X. Z. Wang, Statistical process control charts for batch operations based on independent component analysis. Industrial & Engineering Chemistry Research, 2004, **43**: 6731-6741.

1323

Control of Fed-Batch Yeast Cultivation Using a Capacitance Sensor

Giann B. Reis, Antonio C. L. Horta, Teresa C. Zangirolami, Roberto C. Giordano and Antonio J. G. Cruz

UFSCar - Department of Chemical Engineering - Federal University of São Carlos.

LaDABio – Laboratory for Development and Automation of Bioprocesses.

Rod. Washington Luis (SP 310), Km 235, CEP 13565-905, São Carlos - SP, Brazil.

Abstract

Saccharomyces cerevisiae biomass is still one of the most important biotechnological products in the world, mainly in the form of baker's yeast. Costs of substrates have an important position in the overall economics of baker's yeast production, which is usually run in fed-batch bioreactors. In this context, the maximization of biomass yields, aiming at the highest achievable volumetric productivity is a driving force for improving the control of the process. Classical model-based control strategies for this process, however, face difficulties due to the inherent variability of this system: microorganisms have a complex growth dynamics, lumped in very simplified growth models; raw materials are variable and may be not traceable; the system response depends on non-controlled previous process stages, such as strain selection and inoculum preparation. Hence, the possibility of using feedback information from *in situ* sensors for re-tuning control parameters is an important issue to ensure sub-optimal performances, at least. A capacitance sensor is a device that can monitor cell concentration on-line. In fed-batch cultivations, the feed rate and correlated inputs can be controlled by systems coupled with softsensors that infer the state of the system from on-line measurements of primary variables. Several estimation techniques have been proposed in the literature, and among them the gas balance technique is widely used. The specific respiration rate (qO2), the specific carbon dioxide production rate (qCO2), and the respiratory quotient (RQ) are the main variables determined from the gas balance. Values of RQ for complete oxidation of some carbon sources to carbon dioxide and water are found in the literature. For the baker's yeast production process growing in glucose, a value of RQ close to unity indicates the preponderance of the aerobic route. The capacitance sensor, after calibration, may provide information about cellular growth and viability: the capacitance of the medium is linearly proportional to viable cell concentration. In this work, the signal of capacitance sensor and the RQ value are coupled to a fuzzy algorithm in order to control the glucose feed rate during baker's yeast aerobic cultivation.

Keywords: biomass, process control, *Saccharomyces cerevisiae*, soft-sensor.

1. Introduction

One of the most important variables in bioprocess control is the viable biomass concentration, which can be monitored by capacitance probe (Kiviharju et al. 2007; Xiong, et al., 2008). Until today, the feed rate in fed-batch fermentation was controlled based only in gas balance technique and in simulations results. Some works, as the one

of Xiong et al. (2008), demonstrate that the capacitance probe is a powerful tool for monitoring *Saccharomyces cerevisiae* fermentation, providing a more reliable process control when it is employed.

The capacitance probe functioning principles are shortly described in the followings. At frequencies between 0.1 and 10 MHz, the undamaged cell membrane, *i.e.* a viable cell, acts as an electric capacitor. The dielectric permittivity (ε in pF/cm) is proportional to the viable biomass concentration (Neves et al., 2000). The permittivity and the conductivity (σ in mS/cm) values can be calculated by the Equation 1 and 2 respectively. K (cm^{-1}) is the geometric constant of the electrodes (Markx and Davey, 1999).

$$\varepsilon = C \cdot K \qquad\qquad\qquad\qquad\qquad\qquad (1)$$
$$\sigma = G \cdot K \qquad\qquad\qquad\qquad\qquad\qquad (2)$$

At low frequency, the oscillation of the electric field is slow enough to completely polarize the cells and the medium. When the frequency of the electric field is slow enough, cell polarization reaches a plateau which corresponds to the maximal polarization. At high frequency, the field oscillates quickly and the intracellular charges (ions) do not have time to reach the plasma membrane before the change in the field polarity. Therefore the higher the frequency is, the lower the cell polarization will be. At high frequency, only the rotation of charged molecules induces a permittivity signal. Thus the permittivity value at 10 MHz gives a measurement of the medium's influence only. The difference of the permittivity measurements at low and high frequency is proportional to viable cellular concentration, as illustrated at Figure 1.
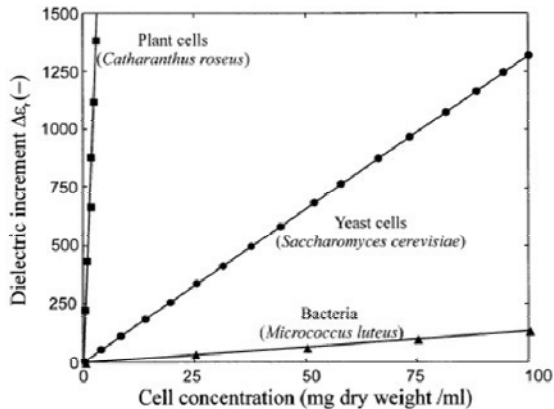
Figure 1. Permittivity (dielectric) increment of the β-dispersion as a function of biomass concentration, for a number of cell types (Markx and Davey, 1999).

2. Materials and Methods

2.1. Experiment

Experiments were carried out in a 6 L stirred tank bioreactor (5 liters working volume), filled with 4.0 liters of medium solution (glucose, 1.2 g/l; KH_2PO_4, 5.0 g/l; $MgSO_4 \cdot 7H_2O$, 0.5 g/l; $(NH_4)_2SO_4$, 4.5 g/l, yeast extract, 3.0 g/l), in fed-batch mode. The inoculum was inserted at concentration of 1 g/l and the cultivation process takes place. Stirrer speed was set to 350 rpm. The level of dissolved oxygen (DO) was maintained

over 45% of saturation (Kasperski and Miskiewicz, 2008). A schematic representation of the experimental set-up used in this study is shown in Figure 2.
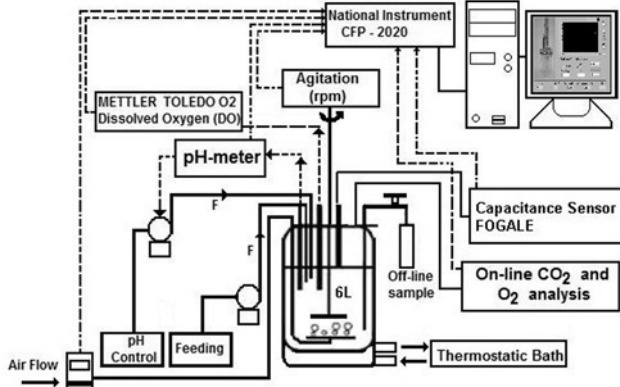
Figure 2- Schematic representation of the experimental system used in this study.

2.2. Off line variables

Dry cell weight (DCW, g/l) was measured by gravimetric method (samples were centrifuged at 8000×g for 15 min followed by drying at 100°C for 24 h). Cell viability was assessed through the Methylene Blue staining method, followed by counting of stained and non-stained cells in Neubauer chamber (Antonini, 2004).

Glucose and ethanol concentrations were measured by HPLC.

2.3. On-line variables

Cell concentration (viable cells) was measured by Biomass System – Viable Cell Density Monitoring (Fogale Nanotech). Equation 3 presents a calibration curve relating viable cell concentration (C_x) and permittivity (ε).

$$C_x(t) = 0.029 \cdot \varepsilon(t) - 3.23 \qquad (3)$$

pH was monitored by pHmeter (GLI PRO P3, Hach, USA). The exhaust gas (oxygen and carbon dioxide molar fraction) was measured by the gas analyzer (SICK-MAIHAK S710, German). The dissolved oxygen concentration was monitored (O_2 4050, Mettler Toledo, Switzerland).

2.3.1. The control algorithm

In this study, a Mamdani configuration for the fuzzy algorithm was used. A basic fuzzy logic control structure consists of a *Fuzzifier* to transform measured data into suitable fuzzy sets; an *Inference engine-fuzzy rule base* to evaluate assigned fuzzy rules and has the capability of simulating human decision making by performing approximate reasoning; a *Defuzzifier* to convert fuzzy sets into a single number.

There are several methods for extracting a crisp value from a dispersion of experimental data to build up a fuzzy set. The center-of-gravity method was adopted in this work for the *defuzzification* step. It should be stressed that the decision on which *defuzzification* method should be applied relies mostly on heuristics and on the experience of the programmer. More details about fuzzy logic can be found in Zadeh (1965), Hisbullah et al. (2003), Tahera (2008) and Karakuzu et al. (2006).

The first step in designing a fuzzy control is to identify fuzzy input and output monitoring variables. During *S. cerevisiae* cell growth, a strong relationship was observed between experimental permittivity and carbon dioxide concentration in the exhaust gas. The fuzzy

logic algorithm presented here has 16 rules (Table 1). It was written in MatLab® and implemented in a LabVIEW® environment. Data strings were exchanged in real time between the MatLab program and the data acquisition system (programmed in LabVIEW).

The variables chosen as input to the fuzzy algorithm were the Respiratory Quotient (RQ) and the error (e), a variable defined as the difference between the reference cellular growth coefficient (μ_{REF}) and its on-line estimated value (μ). The reference value had to be modified during cultivation because the growth behaved differently as the experiment progressed. Its value ranged from 0.06 to 0.1. The output variable was glucose feed flow rate powered by peristaltic pump (Ismatec, Switzerland). The rules of fuzzy logic algorithm included the two input variables, each one associated to four fuzzy sets: a) antecedent: input RQ – Low (RQL), Good (RQG), High (RQH) and Very High (RQVH); input error (e) – Negative Difference (ND), Low Difference (LD), Medium Difference (MD) and High Difference (HD); b) consequent: glucose feed flow rate – Null Flow (NF), Low Flow (LF), Medium Flow (MF) and High Flow (Table 1).

Table 1 – Fuzzy set rules of input and output variables.

Rule	IF (antecedent)	THEN (consequent)
1	RQL **AND** ND	LF
2	RQG **AND** ND	LF
3	RQH **AND** ND	LF
4	RQV **AND** ND	NF
5	RQL **AND** LD	MF
6	RQL **AND** MD	MF
7	RQL **AND** HD	HF
8	RQG **AND** LD	LF
9	RQG **AND** MD	LF
10	RQG **AND** HD	MF
11	RQH **AND** LD	NF
12	RQH **AND** MD	NF
13	RQH **AND** HD	NF
14	RQV **AND** MD	NF
15	RQV **AND** ND	NF
16	RQV **AND** HD	NF

3. Results and Discussion

The correct choice of membership function plays an essential role in achieving a successful design of fuzzy logic controller. Unfortunately, this is not deterministic and has no unique solution. Furthermore, this choice is based on subjective decision criteria and relies heavily on time consuming trial and error. Different membership functions were tested during the fuzzy parameters tuning (*Gaussian*, *Bell*, *Triangular* and *Trapezoidal*) and the best result was obtained using triangular and trapezoidal membership functions, as illustrated in Figure 3. Figure 4 shows the cell, glucose and ethanol concentrations during the experiment.

The exponential growth phase started shortly after the inoculation. Glucose was exhausted within 1 hour of cultivation. Figure 5 shows the behavior of the manipulated variable (glucose feed flow rate) and controlled variable as well (RQ).

At the start of the feed (between 1 and 4.5h), the parameters of the fuzzy controller were changed by the operator, leading to sharp fluctuations of the feeding pump. After the right set of parameters were adjusted, the behavior of the glucose feed flow rate stabilized, as can be seen in Figure 5.

(a)   Input variable – error ($\mu_{REF-}\ \mu$).

(b)   Input variable- RQ.

(c)   Output variable - Glucose feed flow rate.
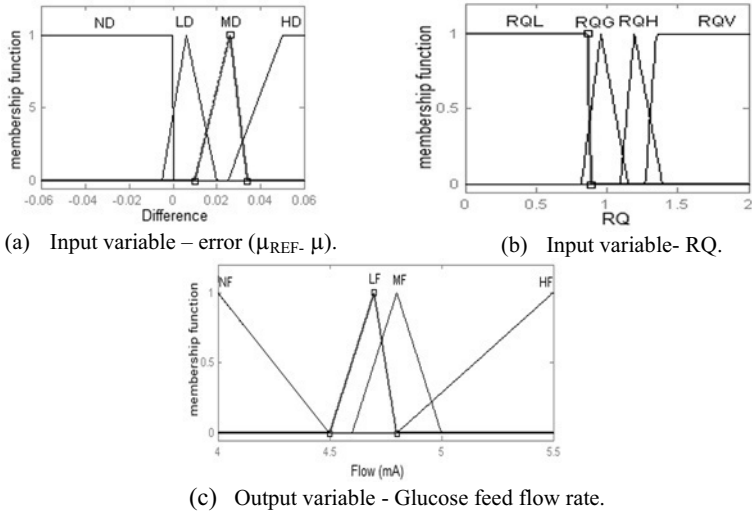
Figure 3- Final membership functions of antecedent after parameters tuning: (a) input variable (μ_{REF} - μ); (b) input variable RQ; (c) consequent part – Glucose feed flow rate.
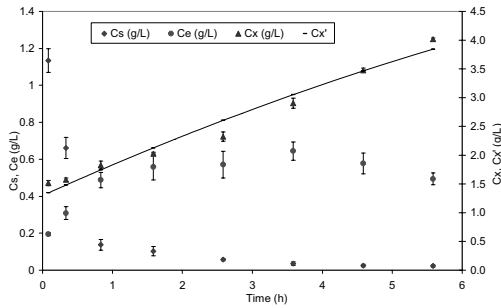
Figure 4 - Glucose concentration (Cs), ethanol concentration (Ce), biomass concentration measured by dry weight (Cx) mass, viable biomass concentration provided by capacitance probe (Cx').
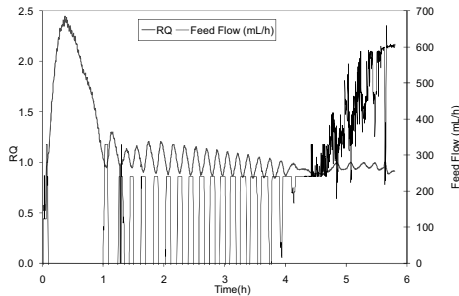
Figure 5 - Behavior of the manipulated variable (glucose feed flow rate) and controlled variable (RQ).

4. Conclusion

The fuzzy control algorithm using the permittivity and the RQ values as input variables was successfully implemented. The glucose concentration into the medium was kept at low values (close to zero) during the whole assay. As a consequence, ethanol concentration was kept low. Cellular growth rate was kept constant until the end of the experiment. The cell viability remained at 98% throughout the experiment. This information allows to correlate directly the probe signal to the cell concentration. The cell productivity ($\Delta Cx/\Delta t \cong 0.45$ g/l/h) reached was considered high if compared to others reported at the litterature, as $\Delta Cx/\Delta t \cong 0.41$ g/l/h obtained by Oliveira (2006). The results demonstrate that the use of the capacitance sensor to reach the optimun control of system is viable, and the fuzzy control was able to control glucose feed flow rate efficiently using this probe signal.

References

Antonini, S.R.C., 2004, Métodos de análises e monitoramento microbiológico em laboratórios de destilaria. Dept. Tecnologia industrial e Sócio-Economia Rural Centro de Ciências Agrárias – Campus Araras.

Hisbullah, M.A.H., Ramachandran, K.B., 2003, Design of a fuzzy logic controller for regulating substrate feed to fed-batch fermentation. *Institution of Chemical Engineers*, 81,138-146.

Karakuzu, C., Turker, M., Özurks., 2006, Modelling, on-line state estimation and fuzzy control of production scale fed-batch baker's yeast fermentation. *Control Engineering Practice*, 14, 959-974.

Kiviharju, K.; Salonen, K.; Moilanen, U.; Meskanen, E.; Leisola, M.; Erikäinen, T., 2007, On-line biomass measurements in bioreactor cultivations: comparison study of two on-line probes. J Ind Microbiol Biotechnol, 34:561–566.

Kasperski, A., Miskiewicz, T., 2008, Optimization of pulsed feeding in a Baker's yeast process with dissolved oxygen concentration as a control parameter. *Biochemical Engineering Journal*. (*In press).*

Markx, G.H.; Davey, C.L., 1999, The dielectric properties of biological cells at radiofrequencies: Applications in biotechnology. Enzyme and Microbial Technology, 25: 161–171.

Neves, A.A., Pereira, D.A., Vieira, L.M., Menezes, J.C., 2000, Real time monitoring biomass concentration in *Streptomyces clavuligerus* with industrial media using a capacitance probe. Journal of Biotechnology, 84: 45-52.

Oliveira, J. A., 2006, Modelagem Matemática do Processo de Leveduras de Panificação – Um Estudo de Caso, dissertation of masters degree, *Federal University of São Carlos – SP- Brazil.*

Tahera, K., Ibrain, R.N., Lochert, P.B., 2008, A fuzzy logic approach for dealing with qualitative quality characteristics of a process. Expert System with Applications, 34, 2630-2638.

Xiong, Z.Q.; Guo, M.J.; Guo, Y.X.; Chu, J.; Zhuang, Y.P.; Zhang, S.L., 2008, Real-Time Viable-Cell Mass Monitoring in High-Cell-Density Fed-Batch Glutathione Fermentation by *Saccharomyces cerevisiae* T65 in Industrial Complex Medium. Journal of Bioscience and bioengineering, Vol. 105, No. 4, 409–413.

Zadeh, L. A., 1965, Fuzzy sets. *Information and Control*, 8, 338–353.

1329

Monitoring of Vinyl Chloride Suspension Polymerization Using NIRS. 2. Proposition of a Scheme to Control Morphological Properties of PVC

João Miguel de Faria Jr.,[a,b] Fabricio Machado,[b,‡] Enrique Luis Lima,[b] José Carlos Pinto[b]

[a]*Braskem S.A., Rua Hidrogênio, 3342, Pólo Petroquímico, CEP: 42810-000, Camaçari, BA, Brazil*
[b]*Programa de Engenharia Química / COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária, CP 68502, Rio de Janeiro, 21945-970, RJ, Brazil*

Abstract

This article introduces the use of in-line and *in situ* monitoring of the dynamic evolution of PVC morphological properties in suspension polymerization reactions based on near infrared spectroscopy (NIRS) intended for control applications. It is shown for the first time that it is possible to follow the dynamic evolution of morphological properties (such as the bulk density, BD, and the cold plasticizer absorption, CPA) in real time. The obtained results indicate that closed-loop control schemes can be properly implemented to control end-use resin properties, which depend upon the morphology of PVC polymer particles. Suspension polymerization of vinyl chloride performed in bench and pilot scale showed that process variables, such as the agitation speed and the amount of suspending agents) can be successfully manipulated in order to control the morphological parameters of PVC resin. Based on this information, a control scheme for these parameters is proposed.

Keywords: Near Infrared Spectroscopy (NIRS), MVC Suspension Polymerization, Dynamic Evolution of Morphological Properties, Control of Polymer Particle Morphology.

1. Introduction

It is well known that profits and plant operation quality may be significantly increased through the implementation of plant automation and the adoption of advanced control strategies. This is particularly true in the polymer industry, as polymer materials are characterized as "product-by-process" materials, which means that the history of the reaction is of great importance to define the final product properties and quality. In the polymer industry it is generally very difficult to minimize specifications drifts caused by process disturbances and uncertainties through blending of different batch products and/or additional separation steps, as usually performed in other fields (Vieira et al., 2002). Specifically, in the case of poly(vinyl chloride) (PVC) resin, a sophisticated arrangement of particles morphology appears as a special feature of PVC resins

---

‡ Current Address: Instituto de Química, Universidade de Brasília, *Campus* Universitário Darcy Ribeiro, CP 04478, 70910-900, Brasília, DF, Brazil, E-mail: fmachado@unb.br

obtained from suspension polymerization process. Very poor solubility of PVC in the vinyl chloride monomer (VCM) causes precipitation of PVC in the VCM droplet at early stages of polymerization, proceeding with the generation of complicated particle structures. This PVC particle structure is very important for understanding of process capability and quantities of produced product (Saeki and Emura, 2002).

The production of polymers with pre-specified properties has placed great emphasis on the development of accurate and robust instruments for the on-line monitoring of polymerization reactions. The complex nature of polymerization systems is one of the main contributing factors to the difficulty of on-line measurements. The presence of long measurement delays and poor reliability are two common problems associated with the on-line characterization of the polymer quality. Even the off-line measurement of many polymer properties is a difficult task, requiring sophisticated and time-consuming analytical techniques. In addition, the development of on-line polymer sensors requires a multidisciplinary effort, including mathematical modeling and data processing, improved knowledge and understanding of the process, reactor design, and modern advanced control and instrumentation engineering (Machado and Bolzan, 1998; Kammona et al., 1999).

An important problem is that, generally, in batch processes, the polymer quality can be assessed only a certain period of time after the end of the reaction. Therefore, many times, the quality of raw material changes (e.g. the quality of monomer) during this period and probably these changes, that could probably affect the resin quality parameters, may not be detected immediately. Consequently, off-specification products can be produced. In addition, as the process will be in a different state, adjustments in the formulation performed based on this delayed information would not assure the solution of polymer quality problems. With the methodology proposed here, it will be possible to track the dynamic evolution of PVC morphological parameters in real time, which means that the long delay times involved in the suspension polymerization of PVC will be properly eliminated, making possible the use of control actions related to the real state of the process.

2. Experimental Procedure

A standard recipe consisting of poly(vinyl alcohol), initiator, water, monomer and additives was fed into the reactor to perform the polymerization reactions. The detailed recipe cannot be presented for proprietary reasons. Detailed description of reactor apparatus is presented by de Faria Jr. et al. (2009), and the reader is referred to this publication for a more detailed description of the protocol.

The studies were divided in three groups. The first one was performed to obtain the dynamic evolution of bulk density (BD) and cold plasticizer absorption (CPA). The second one was performed to verify if changes in the agitation speed would lead to modifications of the BD, CPA and average diameter of final PVC particles. The tests were performed in a bench stainless-steel reactor with a volume of 0.01 m^3 and maximum working pressure and temperature of 20 bar and 250ºC, respectively. A thermocryostatic bath was used to control the reaction temperature. The agitation speed could be varied in the range between 0 and 2400 rpm and a Pt 100 thermocouple was used to monitor and control the reactor temperature.

After 20 minutes of reaction, the normalized agitation speed was changed in the range from 0.40 to 1.00. In all batches the normalized agitation speed at the beginning of the reaction was equal to 0.6. The runs were replicated to analyze the process

Monitoring of Vinyl Chloride Suspension Polymerization Using NIRS. 2. Proposition of a Scheme to Control Morphological Properties of PVC

1331

reproducibility. For each run, BD, CPA and PSD of the final PVC powder were obtained.

The third group of studies was performed by dosing suspending agent into the reaction medium in order to verify if changes of the suspending agent concentration would lead to modifications of the BD, CPA and average diameter of final PVC particles, as in the previous. Tests were performed in a pilot stainless-steel reactor with volume of 0.1 m^3. A cascade control mode (PID-PI) combined with a Distributed Control System was used to control the flow rate of hot water in the jacket. After 20 minutes of reaction, the normalized amount of suspending agent was dosed into the reactor (0.58, 0.75, 0.83, 1.00). Additions were performed in different batches, using a solution of dispersant agent in water. The solution was kept inside a process vessel that was isolated from the reaction medium by a handle valve. After 20 minutes of reaction, the valve was opened, allowing for the solution to flow into the reactor.

3. Results and Discussion

The correlations obtained to predict BD and CPA were excellent and able to represent qualitatively the dynamic evolution of both variables during the polymerization reactions, as shown in Figure 1. The quality of prediction for both BD and CPA, as obtained with the calibration model, can also be regarded as very good. It is very important to emphasize that calibration was performed only with samples obtained at the end of the batch, because quality parameters can only be measured when the polymerization is finished. In spite of that, the dynamic evolution of BD and CPA are predictable and follow very well defined trajectories, showing the robustness of the proposed monitoring scheme (in Figure 1, only the last points of the BD and CPA dynamic trajectories are real experimental points).
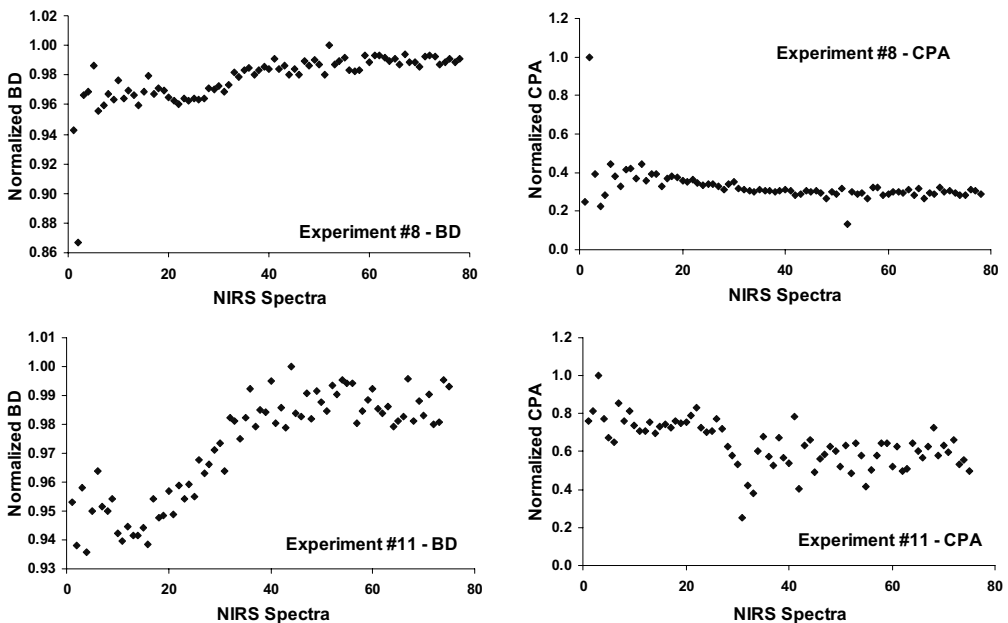
Figure 1 - Dynamic Evolution of BD and CPA during the VCM Polymerization, as predicted by the PLS model.

Figure 1 shows that *in situ* and on-line monitoring of morphological PVC properties using NIRS can be employed successfully in MVC suspension polymerization process, as dynamic trajectories are smooth and well-behaved, indicating that model predictions can be useful along the whole reaction course and used for monitoring and control purposes.

The average diameter of PVC particles obtained at the end of the polymerization reactions was analyzed by screening and light scattering, showing good agreement, as shown in Figure 2. This figure illustrates the effect of agitation speed on the average particle size of PVC. Based on these results, one can conclude that it is possible to vary the agitation speed in order to control the particle size distribution and CPA of PVC particles. As shown, the sudden increase of the agitation speed during the polymerization leads to the diminishment of the average droplet diameter and consequently of the final PVC particle (Dp). After attainment of a minimum Dp value, the overall superficial particle area increases so much that PVC particles agglomerate to reduce the interfacial energy of the system. If stirring continues to increase, the break-up rates due to the high shear forces predominates and Dp decreases again (see in Figure 2 the value of normalized agitation speed equal to 1).
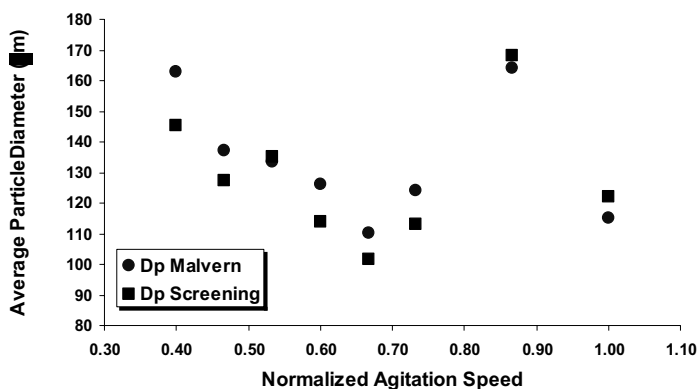
Figure 2 - PVC Particle Diameter Obtained by Screening and Light Scattering.
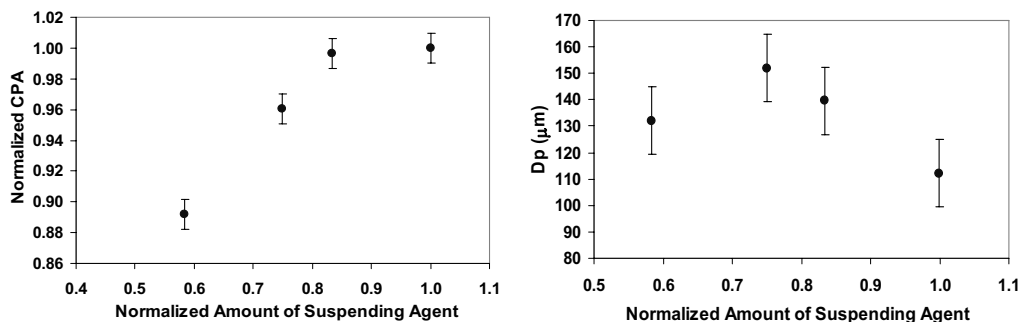
Figure 3 - Effect of Dosing of Suspending Agent on the Bulk Density and Cold Plasticizer Absorption.

According to the results obtained from the third group of experiments, one can conclude that it is possible to dose suspending agent during the reaction to control the particle size distribution, BD and CPA of the final PVC resins, since polymer particles with different diameters, BD and CPA can be obtained when the dispersant agent is dosed

into the reaction medium during the run, as shown in Figures 3. It is important to observe that these experiments were performed in a pilot reactor, where the reactor volume was bigger than the bench reactor volume where the experiments of group two were performed. These observations are in very good agreement with independent observations in the industrial process (although not shown here).

Based on the results presented in Figures 1 to 3, a control scheme is proposed for monitoring and control of important morphological properties of PVC obtained in suspension polymerization process. The scheme takes into account the acquisition of signal available at polymerization site plus the signal from de NIR probe. Among the signals of practical interest available in the polymerization system, one may cite the agitation speed (that provides information about the evolution of the medium viscosity), the reactor pressure (that provides information about the monomer conversion), the reactor temperature (that provides information about the reaction rate, the medium viscosity and the relative volatility of VCM), the inlet and outlet jacket temperature (that allows for implementation of calorimetric schemes for evaluation of monomer conversion, heat transfer coefficients and reaction rates). Figure 4 shows the proposed scheme for the monitoring and control of PVC morphology and the reaction state in real time.
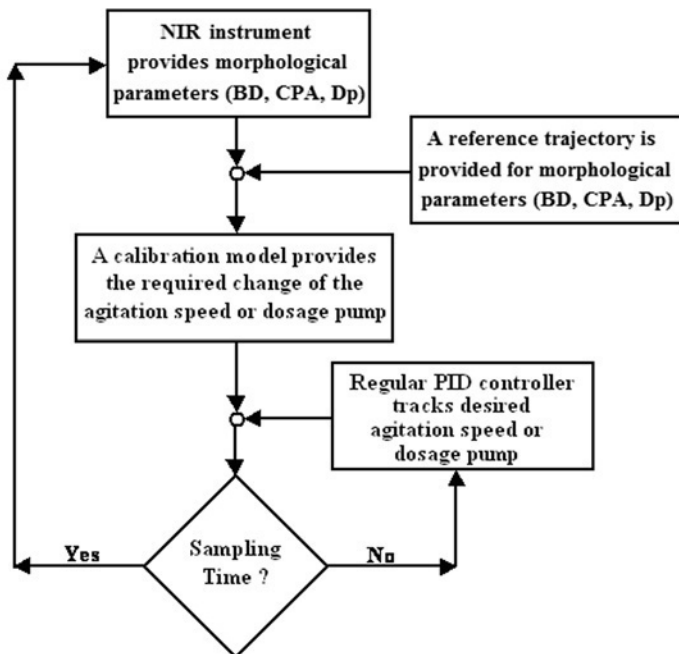
Figure 4 - Scheme for Monitoring and Real Time Control of the Morphological Properties of PVC Particle.

The implementation of the proposed in-line monitoring and control scheme will permit to follow, in real time, the evolution of conversion, of morphological features of PVC particles and the quality of the heat exchange. According to this scenario, the measured trajectories can be compared with ideal trajectories designed off line for control purposes. These ideal trajectories can result from optimization of any variable of interest (for example, minimum reaction time) or from a standard run that can be regarded as a reference one. In the second case, the NIRS signal obtained along a

standard run can be transformed into average size, CPA and/or BD (obtained from calibration model) and used as setpoints for others batch runs. In this case, different setpoints can be used to produce different PVC grades. According to this scheme, the process model can also be used to estimate in real time the monomer conversion and the quality of heat exchange (with the help of a simple balance model).

Reactor temperature should not be manipulated for control of morphological characteristics of PVC particles because it affects the K-value of the resin very significantly. When it is not possible to manipulate the initiator concentration along the polymerization batch, the reaction time can be manipulated through selection of initiator cocktails in the initial charge. Finally, the morphological variables (BD, CPA and average particle size, predicted by NIRS signal) can be controlled through proper manipulation of agitation speed and\or dosing of suspending agent. Thus, whenever a deviation in the variable of interest is detected (for example, Dp), based on the obtained calibration model, a control action must be sent to the agitation speed and/or dosing pump, allowing for satisfaction of the desired trajectories (Dpset). For example, if Dp is bigger than the Dpset, the stirring can be increased and/or the dosing pump can feed an additional amount of suspending agent into the reactor vessel. On the other hand, if Dp is lower than Dpset, the system agitation can be reduced.

4. Conclusions

It was shown for the first time that it is possible to follow the dynamic evolution of morphological properties (such as the bulk density, BD, and the cold plasticizer absorption, CPA) in real time during the PVC suspension polymerization. The obtained results indicated that closed-loop control schemes can be properly implemented to control end-use resin properties, which depend upon the morphology of PVC polymer particles. Suspension polymerization of vinyl chloride performed in bench and pilot scale showed that process variables, such as the agitation speed and the amount of suspending agents) can be successfully manipulated in order to control the morphological parameters of PVC resin. Based on this information, a control scheme for these parameters is proposed.

Acknowledgements

References

J. M. de Faria Jr., F. Machado, E. L. Lima, J. C. Pinto, 2009, Monitoring of Vinyl Chloride Suspension Polymerization Using NIRS. 1. Morphological Properties Prediction, Computer-Aided Chemical Engineering (Accepted).

R. A. F. Machado, A. Bolzan, 1998, Control of batch suspension polymerization reactor, Chemical Engineering Science 70, 1-8.

R. A. M. Vieira, C. Sayer, E. L. Lima, J. C. Pinto, 2002, Closed-Loop Composition and Molecular Weight Control of a Copolymer Latex Using Near-Infrared Spectroscopy, Industrial & Engineering Chemistry Research, 41, 12, 2915-2930.

O. Kammona, E. G. Chatzi, C. Kiparissides, 1999, Recent Developments in Hardware Sensors for the On-Line Monitoring of Polymerization Reactions, Journal of Macromolecular Science - Reviews in Macromolecular Chemistry and and Physics, C39, 1, 57-134.

Y. Saeki, T. Emura, 2002, Technical Progresses for PVC Production, Progress in Polymer Science, 27, 2055-2131.

1335

Modified Unsceted Recursive Nonlinear Dynamic Data Reconciliation for Constrained State Estimation

Sachin C. Kadu,[ab] Mani Bhushan,[a]• R.D. Gudi,[a] Kallol Roy[c]

aDepartment of Chemical Engineering, IIT Bombay, Mumbai 400076, India
bReactor Projects Division, BARC, Mumbai 400085, India
cResearch Reactor Maintenance Division, BARC, Mumbai 400085, India

Abstract

In state estimation problems, often, the true states satisfy certain constraints that need to be incorporated during the estimation procedure. Amongst various constrained nonlinear state estimation algorithms proposed in literature, the unscented recursive nonlinear dynamic data reconciliation (URNDDR) proposed by Vachhani et al. (2006) seems to be promising since it is able to incorporate constraints while maintaining the recursive nature of estimation. In this article, we propose a modified URNDDR algorithm that gives superior performance compared to basic URNDDR. The improvements are obtained via better constraint handling and are demonstrated via a representative case study. **Keywords**: URNDDR, Kalman Filter.

1. Introduction

Kalman filter (KF) based state estimation techniques are widely employed to produce accurate estimates of the true states in the presence of noisy sensor information and an uncertain process model. For most of the real world problems of interest, the system dynamics and observation equations are nonlinear (NL) in nature. Most of the NL estimators proposed in literature are extensions of KF. The most widely used estimator for nonlinear systems is extended KF (EKF). The EKF involves simple linearization of the nonlinear terms in the model so that the traditional linear KF equations can be applied. However, the implementation of EKF requires the Jacobian matrices which may be expensive to compute and are prone to numerical errors, especially for higher order systems. Further, the linear approximation of the system may introduce significant errors in the state estimates, which may even lead these estimates to diverge over time. To overcome these difficulties, Julier et al. (1995) proposed the unscented KF (UKF). It yields performance equivalent to the KF for linear systems, and can be used for nonlinear systems without the linearization steps, which are required by the EKF. The fundamental component of UKF is the unscented transformation (UT), which uses a set of deterministically chosen weighted points (called as sigma points) to parameterize the mean and covariance of a probability distribution. However, the UKF suffers from two drawbacks: (i) the updated state estimates are obtained using a linear estimator (Vachhani et al., 2006), (ii) similar to EKF, UKF also does not incorporate constraints on states during estimation. This can lead to meaningless estimates (such as negative concentrations, etc.) in several applications. The recursive nonlinear dynamic data

---

• Corresponding Author, Email: mbhushan@iitb.ac.in; Tel:+91-22-25767214; Fax:+91-22-25726895

reconciliation (RNDDR) method developed by Vachhani et al. (2005) to incorporate constraints on state estimates, also does not account for the effect of nonlinearity in state and measurement equations as well as the effect of constraints while calculating covariance matrices of the error in the predicted and estimated states. The subsequent work by Vachhani et al. (2006) overcame this limitation by combining merits of RNDDR and UKF to propose the unscented recursive nonlinear dynamic data reconciliation (URNDDR). A rigorous analytic method of incorporating state equality constraints in the KF by projecting the unconstrained KF estimates onto the boundary of the feasible region at each time step was developed by Simon and Chia (2002). But this method is applicable only for equality constraints and requires Jacobian calculation. Recently, Kandepu et al. (2008) proposed a simple procedure of projecting transformed infeasible sigma points onto the boundary of feasible region in order to include state constraints in the UKF algorithm. However, this projection idea was not applicable to the case of nonlinear equality constraints. Due to various limitations of alternative constraint handling algorithms as discussed above, the URNDDR proposed by Vachhani et al. (2006) seems to be promising. The key idea in URNDDR is to replace the update step in UKF by a set of optimization problems that are solved at each sigma point to satisfy the constraints on the states. The updated (filtered) state is then obtained as a weighted combination of the individual updated sigma points. Recently, Teixeira et al. (2008) have proposed various approaches based on UKF and URNDDR combinations to solve constrained, nonlinear state estimation problems. However, URNDDR fails to yield feasible state estimates if the constraints are non-convex. Hence, URNDDR can lead to state estimates which do not satisfy state constraints. In this article we propose a modified URNDDR (MURNDDR) formulation that ensures that the filtered state estimate satisfies the specified constraints irrespective of the nature of the constraints (linear/nonlinear, convex/nonconvex, etc.). The improvements due to the modified approach over the basic URNDDR algorithm are demonstrated on a representative case study. Section 2 presents a brief summary of the URNDDR and MURNDDR approach. The utility of the modifications is demonstrated in Section 3 by applying it to a gas phase reaction simulation case study.

2. Estimation Algorithms

2.1. Unscented Recursive Nonlinear Dynamic Data Reconciliation (Vachhani et al., 2006; Narasimhan and Rengaswamy, 2008)
Consider the following state-space description of the process:

$$x_{k+1} = f(x_k, u_k) + w_k \tag{1}$$

$$y_{k+1} = h(x_{k+1}, u_{k+1}) + v_{k+1} \tag{2}$$

$$\text{s.t.} \quad x_L \le x \le x_U; \quad d(x) \le 0; \quad g(x) = 0 \tag{3}$$

In the above model $x_k \in \mathfrak{R}^n$, $u_k \in \mathfrak{R}^m$ and $y_k \in \mathfrak{R}^r$ are respectively the states, manipulated inputs and the measurements at k^{th} time instant. Process noise $w_k \in \mathfrak{R}^n$ and measurement noise $v_k \in \mathfrak{R}^r$ are assumed to be Gaussian with zero mean and covariance matrices Q_k, R_k and uncorrelated with each other. Functions $f()$ and $h()$ represent the process and the measurement model respectively. x_L and x_U are lower and upper bounds on the state vector and $d(x)$ and $g(x)$ represent inequality and equality constraints. For given estimated state $\hat{x}_k(+)$ and covariance $P_k(+)$, URNDDR can be summarized as:

A) Prediction Step: The UT is used to determine mean and covariance of the propagated states. A set of $(2(n+r)+1)$ sigma points and associated weights (z_i) are calculated from following augmented matrices as:

$$\hat{x}_k^a = \begin{bmatrix} \hat{x}_k(+) \\ y_k \end{bmatrix} \text{ and } P_k^a(+) = \begin{bmatrix} P_k(+) & [0]_{nxr} \\ [0]_{rxn} & R_k \end{bmatrix}$$

$$\chi_{i,k}^a(+) = \begin{cases} \left[\hat{x}_k^a(+)\right] & i = 0 \\ \left[\hat{x}_k^a(+) + \theta_{i,k} S_{i,k}\right] & i = 1, ..., 2(n+r) \end{cases} \tag{4}$$

where, $S_{i,k} = \left[\sqrt{P_k^a(+)}\right]_i, i = 1,...,n+r$ and $S_{i,k} = -\left[\sqrt{P_k^a(+)}\right]_{i-(n+r)}, i = n+r+1,...,2(n+r)$ where $\left[\sqrt{P_k^a(+)}\right]_i$ is the ith column of square root of $P_k^a(+)$ matrix and indicates the direction along which the sigma points are selected. In URNDDR, for better approximation of covariance information in presence of bounds, the sigma points used in propagation step are located asymmetrically and the step size $\theta_{i,k}$ for i^{th} sigma point is chosen as:

$$\theta_{i,k} = min(\sqrt{n+\kappa}, \theta_{1,k}, \theta_{2,k})$$
$$\theta_{1,k} = \min_{j:S_{i,k}^j > 0}(\infty, (x_U^j - \hat{x}_k^j(+))/S_{i,k}^j); \quad \theta_{2,k} = \min_{j:S_{i,k}^j < 0}(\infty, (x_L^j - \hat{x}_k^j(+))/S_{i,k}^j) \tag{5}$$

In the above, $S_{i,k}^j$ is j^{th} entry of ith column at k^{th} time instant. Since sigma points are not symmetric the weights corresponding to each sigma points are selected as:

$$z_i = a\theta_{i,k} + b \tag{6}$$

where, $a = \frac{2\kappa-1}{2(n+r+\kappa)(S_\theta-(2(n+r)+1)(\sqrt{n+r+\kappa}))}, b = \frac{1}{2(n+r+\kappa)} - \frac{2\kappa-1}{2\sqrt{n+r+\kappa}(S_\theta-(2(n+r)+1)(\sqrt{n+r+\kappa}))}, S_\theta = \sum_{i=1}^{2(n+r)} \theta_i$

where κ is a tuning parameter. For any symmetric prior distribution with kurtosis *Ku*, choosing κ such that $0 < n+r+\kappa \le Ku$, leads to mean and covariance predictions that are more accurate than those of EKF (Julier et al., 1995) . The transformed set of sigma points are obtained by propagating each sigma point through process model as:

$$\chi_{i,k+1}(-) = f(\chi_{i,k}(+), u_k) \tag{7}$$

The predicted mean and state prediction error covariance matrix (SPECM) are then computed as:

$$\hat{x}_{k+1}(-) = \sum_{i=0}^{2(n+r)} z_i \chi_{i,k+1}(-) \tag{8}$$

$$P_{k+1}(-) = \sum_{i=0}^{2(n+r)} z_i [\chi_{i,k+1}(-) - \hat{x}_{k+1}(-)][\chi_{i,k+1}(-) - \hat{x}_{k+1}(-)]^T + Q_{k+1} \tag{9}$$

In order to account for the effect of Q_{k+1}, sigma points are regenerated from computed $\hat{x}_{k+1}(-)$ and $P_{k+1}(-)$. The new predicted sigma points and modified weights are computed in the same manner as in Equations (4)-(6). These new predicted sigma points and weights are used for update step.

B) Update Step: The optimization problems are solved to obtain the updated sigma points set. The optimization problem corresponding to the ith sigma point is given by:

$$\min_{\chi_{i,k+1}(+)} \left(\chi_{i,k+1}(+) - \chi_{i,k+1}(-)\right)^T (P_{k+1}(-))^{-1} \left(\chi_{i,k+1}(+) - \chi_{i,k+1}(-)\right) + \tag{10}$$
$$\left(y_{i,k+1} - h(\chi_{i,k+1}(+), u_{k+1})\right)^T (R_{k+1})^{-1} \left(y_{i,k+1} - h(\chi_{i,k+1}(+), u_{k+1})\right)$$

subject to the following constraints

$$x_L \le \chi_{i,k+1}(+) \le x_U; \quad d(\chi_{i,k+1}(+)) \le 0; \quad g(\chi_{i,k+1}(+)) = 0$$

The updated sigma points $\chi_{i,k+1}(+)$ $i = 0,1,...,2(n+r)$, are obtained from the solution of above optimization problems, one for each ith point. The updated state and state estimation error covariance matrix (SEECM) are then calculated as:

$$\hat{x}_{k+1}(+) = \sum_{i=0}^{2(n+r)} z_i \chi_{i,k+1}(+); \quad P_{k+1}(+) = \sum_{i=0}^{2(n+r)} z_i [\chi_{i,k+1}(+) - \hat{x}_{k+1}(+)][\chi_{i,k+1}(+) - \hat{x}_{k+1}(+)]^T \quad (11)$$

It can be seen from the formulation (10) that while the individual updated sigma points will satisfy the state constraints (3), the same cannot be said for the updated state since it is a convex combination (11) of the individual updated sigma points. If the feasible region defined by the constraints is nonconvex, then there is no guarantee that a convex combination of the feasible points will yield another feasible point.

2.2. Modified Unscented Recursive Nonlinear Dynamic Data Reconciliation

In order to account for constraints on individual sigma points as well as on the updated state (which is a convex combination of the individual sigma points), in this work we propose to solve the following single optimization problem for the update step:

$$\min_{\substack{\chi_{i,k+1}(+) \\ i=0,...,2(n+r)}} \sum_{i=0}^{2(n+r)} \left[\begin{array}{c} \left(\chi_{i,k+1}(+) - \chi_{i,k+1}(-)\right)^T (P_{k+1}(-))^{-1} \left(\chi_{i,k+1}(+) - \chi_{i,k+1}(-)\right) + \\ \left(y_{i,k+1} - h(\chi_{i,k+1}(+),u_{k+1})\right)^T (R_{k+1})^{-1} \left(y_{i,k+1} - h(\chi_{i,k+1}(+),u_{k+1})\right) \end{array} \right] \quad (12)$$

$$\text{s.t. } x_L \leq \chi_{i,k+1}(+) \leq x_U; \quad d(\chi_{i,k+1}(+)) \leq 0; \quad g(\chi_{i,k+1}(+)) = 0; \quad i = 0,...,2(n+r)$$

$$x_L \leq \sum_{i=0}^{2(n+r)} z_i \chi_{i,k+1}(+) \leq x_U; \quad d(\sum_{i=0}^{2(n+r)} z_i \chi_{i,k+1}(+)) \leq 0; \quad g(\sum_{i=0}^{2(n+r)} z_i \chi_{i,k+1}(+)) = 0$$

In the above formulation, constraints are imposed on individual updated sigma points as well as on the updated state. The decision variables for this optimization problem are the $2(n+r)+1$ updated sigma point vectors. It is also to be noted that in absence of constraints for linear system models, the updated state as computed by the modified optimization problem will be the same as that given by linear KF (this feature is similar to URNDDR). The advantage of (12) over URNDDR is that we are directly minimizing the error in state estimates subject to the constraints on individual sigma point vectors as well as convex combination of the individual updated sigma points. Hence, modified URNDDR approach is guaranteed to satisfy arbitrary nonlinear and nonconvex constraints on the states. In the proposed approach, at the kth time instant, a single optimization problem with $(2(n+r)+1)n$ decision variables is solved instead of solving $2(n+r)+1$ separate optimization problems each with n decision variable as was the case in URNDDR. Rest of the procedure is the same as discussed in section 2.1. The proposed modification ensures that individual sigma points as well as the updated state estimate satisfy the given constraints irrespective of the nature of the constraints. The utility of the proposed algorithm is now demonstrated by applying it to a simulation gas phase reaction case study.

3. Results and Discussion

The URNDDR methodology as proposed by Vachhani et al. (2006) and modified approach as proposed in this article are applied on the gas phase reaction simulation case study (Haseltine and Rawlings, 2003). It involves an irreversible gas phase reaction: $2A \rightarrow B$, with reaction rate $r = k_e P_A^2$. The partial pressure (P.P.) of species A (P_A) and B (P_B) are the state variables and the measurement is of total pressure (i.e. $P = P_A + P_B$). The discretized nonlinear model for the system is (Haseltine and Rawlings, 2003) given in equation (13) where $x = [P_A \quad P_B]^T$. The initial state ($\hat{x}_0$) and associated covariance matrix (P_0) are taken as $[0.001, 0.045]^T$ and $3600 * I_{2X2}$ (where I is identity matrix) respectively, and other parameters are taken from Haseltine and Rawlings (2003).

$$x_{k+1} = \begin{bmatrix} \frac{x_{k,1}}{2k_e \Delta t x_{k,1}+1} \\ x_{k,2} + \frac{k_e \Delta t x_{k,1}^2}{2k_e \Delta t x_{k,1}+1} \end{bmatrix} + w_k ; \quad y_k = [1 \quad 1]x_k + v_k \tag{13}$$

The number of time samples considered is 150. The states as estimated by URNDDR and MURNDDR approach are presented in Figures 1 and 2 respectively. It can be seen from the figures that both approaches accurately track the true states while satisfying the upper and lower bounds (100 and 0 respectively). To demonstrate the ability of the MURNDDR approach to handle arbitrary nonlinear constraints which cannot be incorporated in basic URNDDR, we introduce an additional nonlinear constraint on state as:

$$193 \le D \le 259 \tag{14}$$

where at k^{th} time point, $D_k = \frac{(\hat{x}_k^1(+)-15)^2}{1} + \frac{(\hat{x}_k^2(+)-15)^2}{4}$. The values of upper and lower bounds in (14) are specified such that the true state satisfies this constraint and hence this constraint should improve the quality of estimation. In basic URNDDR, at the kth time point, 7 (number of sigma points) optimization problems, each involving 2 decision variables (number of states) are solved. For the ith optimization problem, constraint (14) is imposed as:

$$193 \le D_{k,i} \le 259; \text{ where } \quad D_{k,i} = \frac{(\chi_{k,i}^1(+)-15)^2}{1} + \frac{(\chi_{k,i}^2(+)-15)^2}{4} \tag{15}$$

On the other hand, in the MURNDDR approach, at the kth time point, a single optimization problem is solved that involves 14 (i.e. 7x2) decision variables and the constraints (14) are given as:

$$193 \le D_{k,i} \le 259 \; for \; i = 1, 2, ..., 7; \text{ and } \quad 193 \le D_k \le 259 \tag{16}$$

The corresponding performances of URNDDR and MURNDDR at first time point are shown in Figures 3 and 4 respectively. It can be seen from these figures that the state (weighted sigma point) estimated by URNDDR violates the nonlinear constraint whereas the state estimated by the proposed MURNDDR satisfies the specified nonlinear constraint. Note that the URNDDR state estimate violates this constraint even though individual sigma points satisfy it since it was incorporated in the optimization formulation corresponding to the ith sigma point. This is due to the inability of URNDDR to guarantee satisfaction of nonconvex constraints by the updated state. However, MURNDDR suggested in this article is able to handle the specified constraint. The performance of URNDDR and MURNDDR with nonlinear constraints is compared by calculating average mean square error (MSE) in estimated states and the average time required for computation for a time point**.** All the computations were performed in Matlab version 7 running in Windows XP environment on a P4, 2GB RAM machine. The results are shown in Table 1. It can be seen from the Table that the quality of estimates is better for MURNDDR due to its ability to handle arbitrary nonlinear constraints. To summarize, this formulation enables incorporation of any arbitrary nonlinear, non-convex constraints on the states while retaining the recursive feature of URNDDR for nonlinear state estimation. These features, as illustrated by the presented case study, make the modified URNDDR a promising tool for nonlinear constrained state estimation. Implementation of the proposed approach on larger problems is currently under investigation.

Table 1 Comparison of performance of URNDDR and MURNDDR

Description	Average MSE	Time (sec)/iteration
URNDDR	0.0301	0.5
MURNDDR	0.0186	2.5
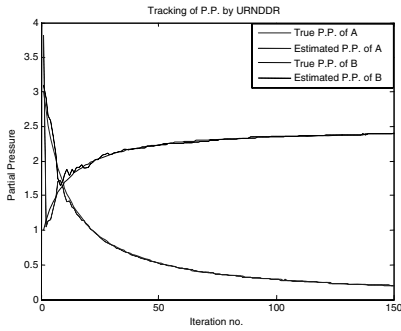
Figure 1: P.P. tracking by URNDDR
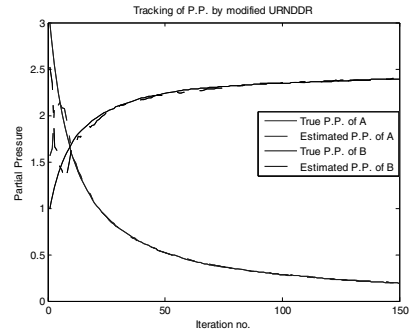
Figure 2: P.P. tracking by MURNDDR
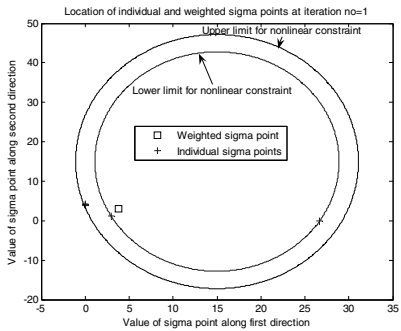
Figure 3: Location of points at first time instance for URNDDR
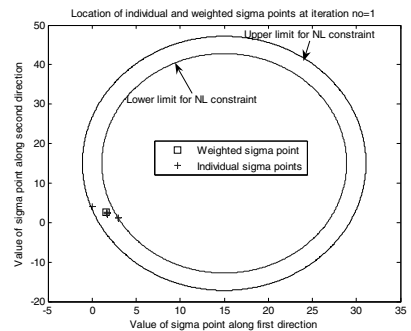
Figure 4: Location of points at first time instance for MURNDDR

References

P. Vachhani, S. Narasimhan, and R. Rengaswamy, 2006, Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation, Journal of Process Control, Vol. 16, pp. 1075–1086.

S. J. Julier, J. K. Uhlmann, and H. C. Durrant-Whyte, 1995, A new approach for filtering nonlinear systems, in Pro. of the American Control Conf., USA, pp. 1628–1632.

P. Vachhani, R. Rengaswamy, V. Gangwal, and S. Narasimhan, 2005, Recursive estimation in constrained nonlinear dynamical systems, AIChE, Vol. 51, No. 3, pp. 946–959.

D. Simon, and T. L. Chia, 2002, Kalman filtering with state equality constraints, IEEE Trans. on Aerospace and Electronic Systems, Vol. 38, No. 1, pp. 128–136.

R. Kandepu, B. Foss, and L. Imsland, 2008, Applying the unscented Kalman filter for nonlinear state estimation, Journal of Process Control, Vol. 18, pp. 753–768.

B. O. S. Teixeira, L.A.B. Torres, L. A. Aguirre, and D. S. Bernstein, 2008, Unscented filtering for interval-constrained nonlinear systems, in Pro. of the 47th IEEE Conf. on decision and control, Cancun, Mexico, pp. 5116–5121.

S. Narasimhan, and R. Rengaswamy, 2008, Reply to Comments on robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation, Journal of Process Control, in press doi:10.1016/j.jprocont.2008.08.002.

E. Haseltine, and J. Rawlings, 2003, A critical evaluation of extended Kalman filtering and moving horizon estimation, Tech. report, TWMCC, University of Wisconsin, Madison, WI.

A Performance Study of Dynamic RTO Applied to a Large-Scale Industrial Continuous Pulping Process

Pablo A Rolandi[a], José A Romagnoli[b]

[a]*Process Systems Enterprise Ltd, London, UK*
[b]*Louisiana State University, Luisiana, USA*

Abstract

A first-principle model-centric approach for advanced control-and-optimisation of industrial manufacturing processes is presented in this work. The elements of this advanced framework are described along with performance studies conducted on industrial control problems using an optimiser based on linear and nonlinear models. The optimal operation of an industrial continuous pulping system is used as a case-study. The software implementation is executed in this dynamic real-time-optimisation (DRTO) environment.

1. Introduction

We envision flexibility and interoperability as the key technological breakthrough of the next generation of model-based APC systems. For example, such APC engine would allow embedding linear models as easily as nonlinear models. Similarly, the APC engine would support empirical/semi-empirical models derived from identification- or reduction-based techniques, as well as fundamental mechanistic models derived from first principles. Concurrently, the APC system would allow unconstrained, quadratic cost problem formulations or general constrained control problems. Finally, this next-generation APC engine would support discrete- and continuous-time formulations interchangeably, and would be integrated into the MPC multivariable control and/or RTO economic optimisation layers of the APC hierarchy. Of course, the form of the optimal control problem does not depend on the characteristics of the APC application and, therefore, a set of mechanisms to formulate (and subsequently interpret) this control problem must be provided. In summary, we argue that the next generation of APC systems should be founded on a domain framework and software platform that allows the interchange of models, solutions methods and control/optimisation settings and strategies seamlessly. Ultimately, such a framework/platform would enable a transparent comparison of controller design and control architecture alternatives.

In this paper, a model-centric platform for advanced process control-and-optimization of industrial processing systems is presented. It provides the appropriate framework through which the aforementioned research issues can be investigated and addressed in a thorough and systematic way. In this work, we present the architecture of the DRTO platform, and we discuss the formulation of industrial optimisation-and-control problems and its subsequent interpretation into a mathematical formalism. The analysis of the controller/optimiser performance using linear and nonlinear models is studied.

2. A Framework for Dynamic Real-Time Optimisation

2.1. Control hierarchy

In this work we devote our attention to the performance of online real-time dynamic optimisation (DRTO) and, therefore, we adopt a control-hierarchy architecture where DRTO directly provides the setpoints of PID controllers (i.e. the regulatory control system), as opposed to the conventional industrial approach where (steady-state) RTO provides the optimum targets to the multivariable MPC controllers.

2.2. Component architecture

The dynamic RTO engine (DRTOE), a software application developed in this work, comprises several software entities: the process-data server (PDS) (which interfaces to OPC servers), the modelling-and-solution engine (MSE), the problem-definition manager (PDM) and the solution-feasibility supervisor (SFS). These are described next.

2.3. Modelling and Solution Engine

The MSE encapsulates all modeling-and-solution services needed by the DRTOE kernel. The advanced process modelling package gPROMS, from Process Systems Enterprise Ltd, has been chosen as the MSE. gPROMS is a state-of-the-art, equation-oriented modelling system that supports hybrid continuous/discrete (HCD) integro-partial-differential-algebraic systems (IPDAEs) of arbitrary complexity. gPROMS meets the requirements of a model server as well as a robust, efficient, general-purpose solution engine.

2.4. Optimiser and virtual plant models

By choosing gPROMS as the MSE, no restrictions are imposed on the type of models that can be embedded in DRTOE. Implicit differential-algebraic systems (DEA) are the most common class of models arising from first-principles' modeling, as well as ordinary differential systems (ODE). In general, these models are highly nonlinear due to natural phenomena such as kinetics, transport and equilibrium (thermodynamics) relationships.

Linear models are a common class of models used in linear constrained/unconstrained APC. Although these models rarely arise from first principles' modelling, they result from linearising nonlinear models around one or more nominal operating points. In the former case, this result in a linear time-invariant (LTI) model; in the latter case, known as successive linearisation, the model is linear time-variant (LTV). gPROMS has built-in linearisation capabilities based on a hybrid structural and numerical algorithm which presents the model in a linear state-space form. LTI models are fully supported by the DRTO engine and LTV models will be included in future work.

A large-scale first-principles' model is available from previous work. This model is used as the virtual plant for all closed-loop runs. In order to avoid plant/model mismatch, this model is also embedded in the control/optimisation engine; naturally plant/model mismatch does exist when this model is subject to linearisation.

2.5. Problem Definition Manager

The industrial operation of continuous and batch processes demand control-and-optimisation problems that change constantly as time progresses. This is due to changes in process conditions and operation specifications. The translation of the control-and-optimisation problem into a high-level dynamic optimization problem is executed by the Problem Definition Supervisor component (PDM). This component interprets the control events posted by users/operators and implements the resulting dynamic-optimisation problem formulation as requested by the MSE. This could be either input

arguments to low-level numerical routines or high-level constructs of declarative modeling languages.

In practice, multivariable, constrained control-and-optimisation problems seeking to improve process performance are difficult to pose and may even result in incorrect or infeasible formulations. Previous studies show that approximately a dozen different types of control events are needed to define a typical industrial control/optimitsation problem. These events have been presented in previous work (Rolandi & Romagnoli, 2008). The so-called "elementary events" have a direct interpretation into the parameterisation of dynamic optimisation problems. For instance, in the case of gPROMS, this means that these events have a direct representation into the gPROMS optimisation entity language. On the other hand, "composite events" are high-level nontrivial instructions that must be reinterpreted into elementary events. Path and zone constraints are two of the most common composite events, and they are supported in the current version of the software. In principle, this concept can be extended to encompass whole transition planning and scheduling scenarios. Overall, the introduction of these composite events greatly simplifies the definition of industrial control-and-optimisation problems by encapsulating operation recipes.

Scaling of decision variables is performed automatically by gPROMS (and most optimisation routines). Scaling of constraints, which is not normally built-in in optimisation modules, is performed by the PDM using the event mechanism described above. This transformation, which is derived from understanding of the optimality criterion, allows fine control of the enforcement of constraints which is needed to improve the robustness and efficiency of industrial control problems.

2.6. Solution and Feasibility Supervisor

The outcome of the PDM is, in general, a constrained dynamic optimisation problem whose solution demands not only optimality but also feasibility considerations. Indeed, the existence of a feasible solution is a key issue to be addressed by any control application. As a consequence, the proposed framework/platform has different build-in strategies to recover from infeasibilities: the Solution-Feasibility Supervisor (SFS) is a dedicated component of the DRTO engine for monitoring the solution and handling of infeasibilities. The SFS is composed by a Solution Interpreter (SI) and a Constraint Manager (CM).

Ideally, the SI would continuously monitor the progress of the open-loop optimisation computation to detect slow convergence between successive iterations due to constraint infeasibility/inconsistency. Due to technical restrictions, the current implementation of the SI logs the magnitude of control variables and constraints violations and overall computation statistics (e.g. number of NLP iterations, line searches and corresponding times) upon termination of the inner optimisation loop, where failing to converge is not considered fatal.

The implementation of the Constraint Manager, a concept that had already been presented in previous publications by these authors (Rolandi & Romagnoli, 2008), is a novelty of this work. The CM is built-in with two independent infeasibility-recovery strategies: i) relaxation vs elimination and ii) ranking vs identification. These individual strategies can be combined to create four different constraint-management policies.

The most common recovery mechanism in commercial MPC packages is constraint ranking and elimination (Qin & Badgwell, 2003). Here, one or more constraints are associated to a ranking level and removed from the control problem formulation upon infeasibility when their priority is below a cut-off level. We have found that this mechanism is sub-optimal and, furthermore, does not guarantee immediate recovery,

leading to inefficient computations. Constraint identification, on the contrary, uses information on the constraints' bounds and values (or their Lagrange multipliers, if available) to detect the subset that is active at any given iteration. Constraints that are violated are either eliminated or relaxed. The relaxation takes place by applying a factor which is provided by the user/operator via a console event mechanism (by default, it is set to 1% of the nominal variable value). Identification is coupled with ranking information because only constraints below a cut-off level can be relaxed or eliminated.

3. Case-study and analysis of results

3.1. Industrial wood pulping
The case study presented in this section is based on an industrial continuous pulping digester (and its auxiliary units) of a state-of-the-art pulp and paper mill. Papermaking is a highly competitive industry where meeting quality control parameters and the demand of highly dynamic markets is paramount to the economics of the business; also, new plants with more flexible and efficient productions schemes often replace old production systems. The key unit in this process is a vertical tubular reactor with several fluid-phase extraction and introduction points along the length of the reactor. The reactor is a solid-liquid heterogeneous system with inter- and intra-particular transport phenomena, where the column of wood chips moves continuously down and the fluid phase moves in co-current or counter-current flow at different heights. Besides the tubular reactor, there is auxiliary heating and heat-recovery equipment consisting of nine unit operations (condensers, heaters, pre-heaters and a kettle-reboiler), and seven transportation and handling devices (the so-called feed line). There are approximately 50 PID control loops, switched to manual or automatic operating mode.

3.2. 3.2 First principles model
Large-scale mechanistic models (that is, models of order ~1E5 variables or so) have seldom been used in advanced model-based control systems, with only a few examples resulting from academic studies rather than industrial applications (e.g. Wisnewski & Doyle, 2001). In this work, a mechanistic model of the continuous pulping system described above has been implemented in gPROMS, resulting in a system of approximately 14,000 algebraic, 1,000 differential equations, and 300 degrees-of-freedom.

3.3. Other implementation aspects
In this work, real-time dynamic optimisation provides the setpoints of key loops of the regulatory control system, condensing the two intermediate optimisation/control layers of a conventional hierarchical control structure (i.e. MPC and RTO) into a single, consistent model-centric application (DRTO).

Depending on the configuration, either this full nonlinear model is embedded in the optimizer or a reduced-order linear model. The latter is obtained by exact linearisation and minimal realisation of the former (the first-principles' nonlinear model) at the nominal steady-state starting point (in both case studies this is at 600.0 ad.ton/day); this is computed by the MSE (gPROMS). Linearisation requires full knowledge of the operating point, that is, degrees-of-freedom and initial conditions of the system. In order to satisfy this requirement as well as the need for state feedback during the movement of the receding horizon, a perfect observer is implemented. This is achieved by accessing the values of states and inputs of the dynamic model used as virtual plant, which is implemented via gPROMS' FOI/FPI communication mechanisms.

While a perfect observer is not feasible in an industrial setting, this choice allows us to focus on algorithmic and performance aspects and avoid any issues introduced by the ubiquitous plant/model mismatch that would characterise any real industrial application.

3.4. Case-study 1

In this case study the main target is to control a production-rate transition from 600.0 ad.ton/day to 650.0 ad.ton/day. From an operations-and-control perspective, the goal is to maximise the pulp yield (Y) at a given production target (P) while maintaining the deviation of pulp selectivity from its quality-control target below a given threshold. Selectivity is measured in terms of the "kappa number" and it is an indication of the proportion of non-cellulosic components present in the wood pulp. The process is initially at steady-state and control actions start taking place three hours before the scheduled production-rate change. The manipulated variables (MVs) optimized by DRTO are the setpoints of three key PID controllers: the chip meter (CM) speed (feed rate of wood chips), and the temperature of the lower (LH) and wash (WH) circulation heaters (which perform indirect heating of the chip column). Two interior-point and two end-point constraints are imposed on the trajectories and final values of selectivity and pulp production rate, and one path constraint is introduced for the deviation of selectivity with respect to its control target. The prediction and control horizons are set to 7 and 5 hr, respectively, and the control window is 1 hr; these settings are representative of industrial practices.
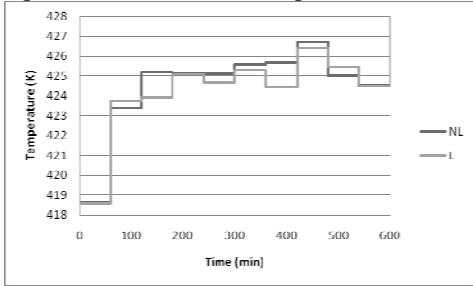
Figure 1: Wash heater (WH), MV.
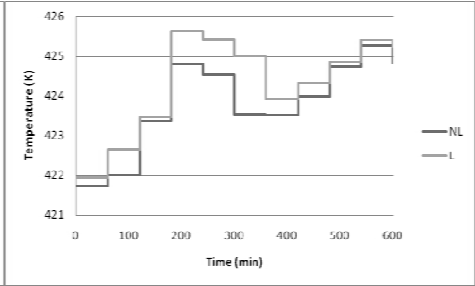
Figure 2: Lower heater (LH), MV.
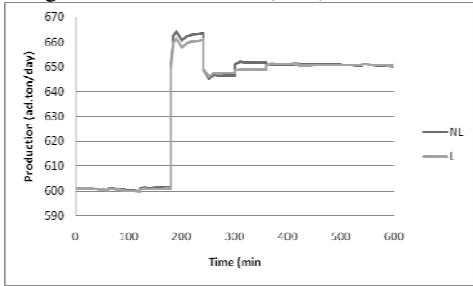
Figure 3: Production (P), CV.
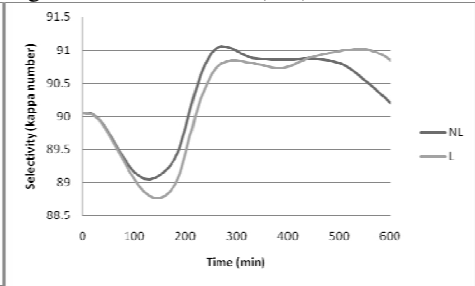
Figure 4: Selectivity (S), CV.

Figures 1 to 4 show the trajectories of key process variables for this case-study, for the linear (L) and nonlinear (NL) embedded model. While the progression of control moves (MVs) is similar for both controllers, the L-controller imposes a higher temperature setpoint for the LH PID loop and a lower setpoint for the WH PID than the NL-controller does. In general, the selectivity trajectories are within the control limits (between kappa number 89 and 91), although the L-controller slightly violates this constraint between 100min and 200min approximately. Overall, the NL-controller and the L-controller exhibit similar control performance. The good performance of the L-

controller is due to fact that the plant is reasonably linear in this limited range of operation and an accurate model of the plant has been obtained by exact linearisartion of the first principles' nonlinear model. Also, the prefect observer periodically corrects the effects of plant/model mismatch by resetting the linear model to the true operating point at every control window. While both controllers execute roughly the same number of NLP iterations, the L-controller is approximately 5 times faster than the NL-controller. Finally, the L-controller hits one infeasibility and recovers, while the NL-controller always remains feasible.

3.5. Case-study 2

This case study is a variation of case study 1. Again, a production-rate transition from 600.0 ad.ton/day to 650.0 ad.ton/day is scheduled, this time, two hours in advance. However, two hours after the transition a failure in downstream processing equipment triggers a production slow-down to the original rate of 600.0 ad.ton/day, which is enforced two hours later. This is a common yet challenging operating scenario in integrated pulp and paper mills which have low inventories by design.

See figures 5 and 6. This time, the NL-controller outperforms the L-controller, since it is able to maintain the constraint on selectivity at all times. The NL-controller also gives a higher mean yield (objective function) throughout the operation horizon, which is an indication that optimality has been achieved (at least locally).
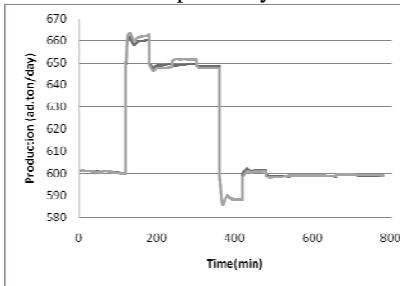
Figure 5: Production (P), CV.   Figure 6: Selectivity (S), CV.
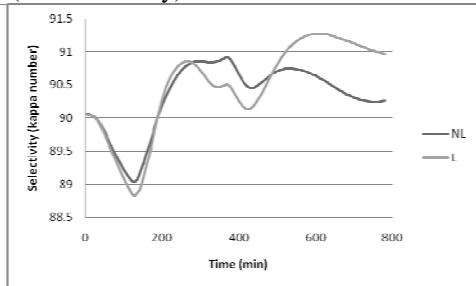
4. Conclusions and Future Work

In this manuscript we presented a framework and software platform that enables the adoption of different model types, solutions methods, and control-and-optimisation strategies transparently and effectively. We examined the performance of a novel DRTO engine through two industrial case studies using linear and nonlinear models. Both first principles model-based controllers performed well and were able to drive the plant through production transitions satisfactorily; while the nonlinear model handled constraints better, the linear model resulted in much faster computational times. The constraint identification and relaxation mechanism for recovery of infeasibilities performed well. Future work will involve refinements to this framework as well as the development of observers for realistic closed-loop feedback with plant/model mismatch.

5. Acknowledgements

References

Wisnewski, P.A. and Doyle, F.J. (2001). IEEE Trans Control Sys Tech, 9, 435-444.
Rolandi P.A., Romagnoli, J.A. (2008). ESCAPE-18, Lyon, France.
Qin, S.J., Badgwell, T.A. (2003). Control Engineering Practice, 11, 733–764.

Control Structure Selection Based Upon Rigorous Dynamic Process Models

Le Chi Pham[a], Sebastian Engell[a]

[a]Process Dynamics and Operations Group, Department of Biochemical and Chemical
Engineering Technische Universität Dortmund
Emil-Figge-Str. 70, 44221 Dortmund, Germany
E-mail address: le-chi.pham | s.engell@bci.tu-dortmund.de

Abstract

In this contribution, the selection of control structures based upon rigorous stationary process models is extended to include the dynamic performance in a consistent manner. A key problem in the comparison of control structures with respect to dynamic performance is that the performance depends on the type and the parameterization of the controllers that are used. In our approach, linear MPC-controllers are assumed and the weights of these controllers are optimized for each structure to yield an optimal economic performance for the disturbance scenarios considered. Thus the result is a good approximation of the best performance that is attainable for each structure and the structures are compared on equal grounds. The approach is demonstrated for a distillation problem.

Keywords: Control structure selection, economic performance, distillation control.

1. Introduction

Control structure selection deals with the choice of manipulated and measured variables used in feedback control. The result of it often has a stronger influence on the performance of the resulting closed-loop system than the choice of control algorithms. Most of the published work from the control community on control structure selection focuses on the dynamic tracking and regulation performance rather than the resulting economic performance of plant, which however should be the main goal of control design for chemical processing plants (see e.g. Engell, 2007). Morari et al. (1980) discussed several important aspects of the control of chemical processes from the point of view of plant performance. This was followed by Skogestad (2000) with the idea of "self-optimizing" control which means that a close-to-optimal steady state performance of the process in the presence of disturbances is established by a well chosen linear control structure. The focus is on the resulting performance at the steady state, and a stepwise approach is proposed to determine the most promising structures. Engell et al. (2005) extended this idea by some refinements replacing the informed judgement by objective criteria and optimization procedures, in particular considering the effect of measurement errors. After the static analysis, the remaining structures were screened by using dynamic performance indicators from linear control theory (Engell et al, 2004).

In the work mentioned above, the disturbances are assumed to be slowly time-varying or constant, so the steady-state behavior is dominant in the economic performance. An incorrect judgment of the performance of the controlled plant may result because the reaction to fast disturbances is not included. In this paper, we extend the procedure

proposed in (Engell et al, 2005, Scharf, 2007) by an evaluation of the dynamic closed-loop performance based upon full rigorous process models.

2. Plant Performance and Control Structure Selection

From a process engineering point of view, the role of automatic feedback control is to keep the process operation as close as possible to the economic optimum in the presence of disturbances and differences between the plant model used for the design of the plant and the real plant. The goal is to regulate the controlled variables to their setpoints in the presence of disturbances in such a way that the optimal manipulated variables u_{opt} are approximated. The effect of feedback control on the profit function J can be expressed as (Engell et al., 2005, Engell, 2007):

$$\Delta J = \left(J(\underline{u}_{nom}, 0) - J(\underline{u}_{nom}, \underline{d}_i) \right) + \left(J(\underline{u}_{nom}, \underline{d}_i) - J(\underline{u}_{opt}, \underline{d}_i) \right) + \left(J(\underline{u}_{opt}, \underline{d}_i) - J(\underline{u}_{con}, \underline{d}_i) \right)$$

The first term is the loss if the manipulated variables are fixed at their nominal values when disturbances occur, the second term is the effect of the optimal adaptation of the manipulated variables in the presence of disturbances, and the third term is the difference between the optimal adaptation and the one which is realized by the chosen feedback control structure. If the first term is large compared with the other terms then changing the manipulated variables will bring little profit, neither online optimizing nor feedback control is useful here. On the other side if the third term is large then online optimizing control (Engell, 2007) should be used instead of feedback control. The overall static performance of a control structure can be quantified by the expected loss of profit (Engell et al., 2005):

$$\Delta J = \int_{-d_{1,\max}}^{d_{1,\max}} \dots \int_{-d_{n,\max}}^{d_{n,\max}} w(\underline{d})(J(\underline{u}_{nom}, \underline{d}_i) - J(\underline{u}_{con}, \underline{d}_i)) dd_1 \dots dd_n$$

where $w(\underline{d})$ is the probability of the occurrence of the disturbance $\underline{d}$. Practically this expression is approximated by the weighted sum over a set of disturbance scenarios.

There are two kinds of disturbances which affect the process: slow and fast varying disturbances. The effect of the former ones can be taken by into account by using a static optimization which analyzes the worst case performance of the regulatory control keeping the controlled variables within a range around the set-points defined by the measurement errors (Engell et al., 2005, Scharf, 2007). How to include the effect of the latter ones is discussed in this paper.

It is proposed to include the dynamic performance in a consistent manner by extending the above analysis to incorporate a full dynamic process model, dynamic controllers, and fast disturbances. In order to avoid the assumption of simple linear control structures and to be reasonably close to the best possible performance that a control structure can yield without excessive numerical effort, linear MPC controllers are assumed As the focus is on the economic performance, the weights used in the cost functions of the MPC-controllers are optimized for each structure with respect to the optimal economic performance for all disturbance scenarios considered.

3. Control Structure Selection Procedure

The proposed control structure selection procedure consists of 6 steps:

3.1. Define the optimization problem:
Analyze the available degrees of freedom of the process during plant operation and choose the manipulated variables. Formulate a profit function *J* to be maximized and specify the constraints that have to be satisfied during the process operation.

$$\max_{u} J(\underline{x}, \underline{u}, \underline{d}_i)$$

$s.t$ system model

$$\underline{h}(\underline{x}, \underline{u}) \le 0$$

The system model can be either dynamic or static, $\underline{h}(\underline{x}, \underline{u})$ are the process constraints, d_i are the disturbances. The output mapping is given by: $\underline{y} = \underline{m}(\underline{x})$, in steady state $\underline{y} = \underline{M}(\underline{u}, \underline{d}_i)$.

3.2. Choose the disturbances:

It is distinguished between two types of disturbances: measurement errors and external disturbances. The former ones can be obtained from the instrument data-sheets. The latter ones may be caused by errors in the assumed model, disturbances, etc.

3.3. Pre-selection of the control structures

The number of possible control structures is given by:

$$C_n^k = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

with n: the number of available measurements and k: the number of controlled variables. This number increases rapidly with the number of measurements. Pre-screening of the unpromising structures is useful here especially for large problems. Several indices can be used, e.g. sensitivity to measurement errors (Engell et al., 2004) or linear performance indicators (Engell et al., 2004).

3.4. Selection of the set-points for regulatory control

The set-points are determined by optimization over the set of disturbance scenarios to fully utilize the potential of feedback control. The optimal set-points are found by solving:

$$\max_{\underline{y}_{set}} \sum_{i=1}^{n} J(\underline{x}, \underline{u}_i, \underline{d}_i)$$

$s.t.: \forall \underline{d}_i :$

$$\dot{\underline{x}} = f(\underline{x}, \underline{u}_i, \underline{d}_i) = 0$$

$$\underline{h}(\underline{x}, \underline{u}) \le 0,\ \underline{y}_{set} = \underline{m}(\underline{x})$$

$$u_{min} \le u_i \le u_{max},\ x_{min} \le x \le x_{max}$$

If the above optimization problem is infeasible, it means that there is no common set-point which can be attained for all disturbances and the given constraints.

3.5. Quantitative evaluation of the benefits of the control structures with constant disturbances

For all scenarios of disturbances $\underline{d}_i$, the worst case control performance for regulation of the controlled variables to values in the range around the nominal set-point $\underline{y}_{set}$ defined by the measurement error e_{sensor} is obtained by solving the following optimization problem:

$$\min J(\underline{x}, \underline{u}_i, \underline{d}_i)$$

$s.t.: \dot{\underline{x}} = f(\underline{x}, \underline{u}_i, \underline{d}_i) = 0$

$$\underline{h}(\underline{x}, \underline{u}) \le 0,\ \underline{y} = \underline{m}(\underline{x})$$

$$\underline{y}_{set} - \underline{e}_{sensor} \leq \underline{y} \leq \underline{y}_{set} + \underline{e}_{sensor}$$

A comparatively large value of the maximum loss means that the corresponding control structure is not able to avoid a poor stationary performance in the presence of the measurement errors and should be excluded.

3.6. Quantitative evaluation of the benefits of the control structures with dynamic disturbances

After the controlled variables that are regulated to their set-points have been determined, the maximum attainable dynamic performance of the controlled system is compared. In order to accomplish this, a simulation of a linear MPC controller which is based upon a linearized model of the plant is employed. The objective function of the MPC controller is defined as

$$P(t_k) = \min_{\mathbf{u}} \left(\int_{t=t_k}^{t_k + H_P} \left(\left\| \Delta y(t) \right\|_P + \left\| \Delta u(t) \right\|_Q \right) dt \right)$$

$$s.t.: \quad \Delta \dot{x}(t) = A\Delta x(t) + B\Delta u(t) + C\Delta d(t) \tag{P1}$$

$$\Delta y = D\Delta x(t)$$

where $\left\| . \right\|_X$ denotes the norm which is defined by: $\left\| u \right\|_X = u^T X u$, X is a positive semi-definite matrix. A, B, C, D are matrices which are obtained by linearizing the system at the operating points found in step 4. P and Q are degrees of freedom and should be chosen such that the economic profit function J is maximized. The prediction horizon H_p is chosen long enough to capture all effects of the disturbances. An upper layer optimization is used in order to compute P and Q over all disturbances:

$$\max_{P,Q} \int_{t=0}^{t=t_{end}} \sum_{i=1}^{n} J(\underline{x}, \underline{u}_i, \underline{d}_i) dt$$

$$s.t.: P, Q \succ 0$$

$$(P1)$$

Thus the result is a good approximation of the attainable performance for each structure. The structure which yields the best performance using the same criterion as step 5 will be chosen.

4. Case study

The methodology described above is applied to a continuous distillation column shown in Figure 1. The distillation column is used for the separation of a binary mixture of Methanol and n-Propanol. The model of the process (Diehl, 2001) is based on the following assumptions: total condenser, negligible vapour holdup, variable liquid holdup, liquid outflow determined by Francis weir formula, constant pressure losses, perfect mixing, the mixture is at equilibrium temperature, Murphree efficiency is applied for each tray. The system model for 40 trays is large and stiff with 82 differential and 122 algebraic variables. There are two operating degrees of freedom after closing the inventory loops: the reflux rate and the heat supply. A composition measurement is assumed to be too expensive so the controlled variables are two tray temperatures. There are: $C_{40}^2 = 780$ possible control structures. The profit function is:

$$J = c_{Methanol} \tanh(x_{Methanol} - desiredvalue)\dot{n}_{Methanol} - c_{heat} heatinput$$

tanh implies that the purity of distillate product should satisfy the specification which is 0.99 in our case. The disturbances are chosen to be: a step change in the feed flow rate

by ±4 l/h, a step change in the feed concentration by ±0.1, a step change in the feed temperature by ±5 K for both the steady state and the dynamic analysis. For the steady state case, additional disturbances are considered: a change in the condenser temperature by ±5 K, a change in the heat loss by ±0.2 kW. In the dynamic scenario, the disturbances are assumed to be uniformly distributed within the lower and upper bound.

The sensor error is assumed to be ±0.33 K. The heat input and the reflux flow rate are restricted to 5 kW and 7 l/h.

First a pre-screening is performed to take unpromising structures out of the large number of options. The generalized criterion of non square RGA (Skogestad and Postlethwaite, 1996) is used. 429 structures with a sum of the elements in row of the RGA less than 0.1 are discarded. Next, a measurement error sensitivity analysis is performed and 78 structures are left for the next step. In step 4, all the structures have a common set-point so 78 structures are considered further in step 5. Only those structures which lead to a profit larger than 85% of the nominal operating profit of 114.893 in the worst case are retained in the final step. 12 structures that satisfy this criterion are compared in the final step with respect to their dynamic performance using the linear MPC simulation with the nonlinear rigorous process model. For the optimization of the weights of the MPC controllers, a gradient-free optimization procedure was used (Berghen and Bersini, 2005). The performance indices of several structures are given in Table 1. Structure (10,31) yields the best performance.
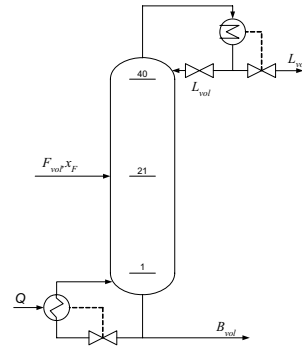
Fig. 1. Distillation Column

Structure	Profit
(6,35)	107.67
(7,33)	106.36
(10,31)	111.02
(11,32)	109.85

Table 1. Performance indices of several structures resulting from a simulation of optimized linear MPC controllers

The results of the MPC simulation of this structure and of the structure (6,35) with disturbances of the feed flow and of the feed temperature are shown in Fig. 2 and Fig. 3. It can be seen that in the presence of strong disturbances in the feed flow by 30% of the nominal flow, the controller with the structure (10,31) manages to keep the controlled variables at their set-points and the top product purity still satisfies the requirement. The structure (6,35) could not keep the purity at the specified value. In the case of a disturbance in the flow rate, the purity drops down close to 0.985. This result verifies the heuristics that the best positions for controlling the composition of a distillation column using two tray temperatures are somewhere in the middle of the rectifying and stripping sections.

5. Conclusion

A methodology for control structure selection was presented with the aim of optimizing the plant performance taking into account the presence of both steady-state and dynamic disturbances. The method was applied successfully to the example of a distillation column. Because of the large computation times for NMPC simulations and in view of the industrial practice to use linear MPC controllers, the performance evaluation in the dynamic case is based upon linear MPC controllers with weights that are optimized for economic performance in an outer loop. Steps 1 – 6 of the control structure selection procedure yield a set of control structures that lead to good performance with respect to

Fig. 2.a Simulation of linear MPC for structure (10,31) with disturbance in the feed flow rate
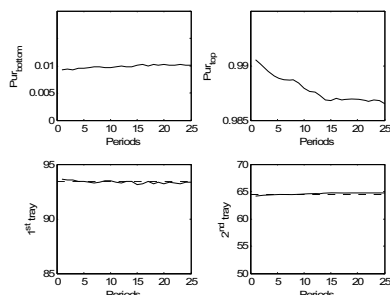
Fig. 2.b Simulation of linear MPC for structure (6,35) with disturbance in the feed flow rate
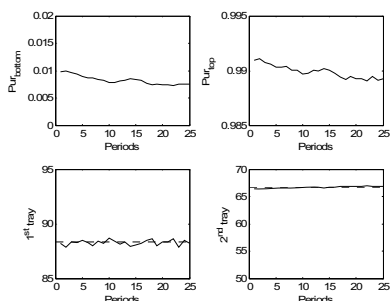
Fig. 3.a Simulation of linear MPC for structure (10,31) with disturbance in the feed temperature
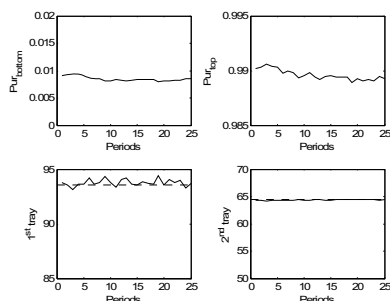
Fig. 3.b Simulation of linear MPC for structure (6,35) with disturbance in the feed temperature

the economic performance of the plant both for stationary or slowly-varying and for fast-varying disturbances. However, it may happen that some of these structures are not suitable for dynamic operation in the presence plant-model mismatch (lack of robustness). To check this, a linear dynamic controllability analysis should also be performed, cf. Engell at al. (2004).

References

F.V. Berghen, H. Bersini, 2005, CONDOR, a new parallel, constrained extension of Powell's UOBYQA algorithm. Experimental results and comparison with the DFO algorithm. Journal of Computational and Applied Mathematics, 157-175.

M. Diehl, 2001, Real-time Optimization for Large Scale Nonlinear Process, Dr. rer. nat. thesis, Universität Heidelberg.

M. Morari, Y. Arkun, G. Stephanopoulos, 1980, Studies in the Synthesis of Control Structures for Chemical Processes, AIChE Journal 26, 220-260.

S. Engell, J.O. Trierweiler, S. Pegel, M. Völker, 2004, Tools and indices for dynamic controllability assessment and control structure selection. In: Integration of Design and Control (P.Seferlis, M.C. Georgiadis, Eds.), 430-464, Elsevier.

S. Engell, T. Scharf, M. Völker, 2005, A methodology for control structure selection based on a rigorous process model. Proc. 16th IFAC World Congress, Prague, paper Tu-E14-TO/6.

S. Engell, 2007, Feedback control for optimal process operation. Journal of Process Control 17, 203-219.

S. Skogestad and I. Postlethwaite, 1996, Multivariable Feedback Control: Analysis and Design, John Wiley and Sons.

S. Skogestad, 2000, Plantwide control: the search for the self-optimizing control structure. Journal of Process Control 10, 487-507.

T. Scharf, 2006, Prozessbezogene optimierungsbasierte Regelungsstrukturauswahl mit Anwendung auf die Reaktiverektifikation, Dr-Ing thesis, Fachbereich BCI, Universität Dortmund (Shaker Verlag, Aachen, 2007).

Model-Based Fault-Tolerant Control of Particulate Processes: Handling Uncertainty, Constraints and Measurement Limitations

Arthi Giridhar[a], Sathyendra Ghantasala[a], Nael H. El-Farra[a]

[a]*Department of Chemical Engineering and Materials Science, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA*

Abstract

This paper presents a methodology for the integrated synthesis of a fault detection and isolation (FDI) and fault-tolerant control (FTC) system for particulate processes described by population balance models. The approach is based on a suitable low-order model of the process and accounts explicitly for model uncertainty, control constraints and measurement errors. The main idea is to shape the fault-free closed-loop dynamics via robust control in a way that enables the derivation of FDI and reconfiguration criteria that are less sensitive to the uncertainties and measurement errors. The results are illustrated through an application to a continuous crystallizer with a fines trap.

Keywords: Robust control, Process monitoring, FDI, Reconfiguration, Crystallizer.

1. Introduction

Particulate processes characterized by the co-presence of continuous and particulate phases are encountered in a large number of processing industries including agricultural, chemical, food, minerals, and pharmaceuticals. Despite the significant and growing body of research work on particulate process control (e.g., see Semino and Ray (1995), Christofides (2002), Zhang and Rohani (2003), Park et al. (2004), Larson et al. (2006) for some results and references in this area), the problem of designing model-based fault diagnosis and fault-tolerant control systems for particulate processes has received limited attention. This is an important problem given the fact that the erosion of controller authority caused by control system malfunctions can directly impact the end product quality and lead to substantial production losses (El-Farra and Giridhar, 2008). It is also a challenging problem due to the infinite-dimensional nature of particulate process models, which precludes their direct usage for control and monitoring purposes, the presence of model uncertainty, control constraints and the inherent limitations on the number and accuracy of process measurements, which can seriously erode the diagnostic and fault-tolerance capabilities of the control system if not explicitly accounted for at the controller design stage. In this paper, we present a robust fault diagnosis and fault-tolerant control (FTC) structure for particulate processes with control actuator faults. Unlike our previous work in El-Farra and Giridhar (2008), the structure presented here addresses explicitly the problems of time-varying model uncertainty and measurement errors. Initially, a finite-dimensional system that captures the dominant process dynamics is derived and decomposed into a number of interconnected subsystems that are each excited by a single manipulated input. A bounded Lyapunov-based controller that satisfies the control constraints is then

designed for each subsystem, leading to an explicit characterization of (1) the fault-free behavior of each subsystem in terms of a time-varying bound on the dissipation rate of the corresponding Lyapunov function, and (2) the robust stability region where constraint satisfaction and robustness with respect to the uncertainty and measurement errors are guaranteed. Using the fault-free bounds as alarm thresholds for fault detection and isolation (FDI), faults in a given actuator are diagnosed by monitoring the evolution of the process within the stability region and declaring a fault when the corresponding threshold is breached. The threshold depends on the achievable degree of asymptotic uncertainty attenuation and thus can be adjusted by properly tuning the controller. To ensure robustness of the FDI scheme to measurement errors, the monitoring region is confined within a subset of the stability region, and the FDI thresholds are enlarged appropriately. The robust FDI scheme is integrated with a controller reconfiguration strategy that orchestrates stabilizing transitions from the faulty actuators to a well-functioning fallback configuration following FDI. Practical implementation issues, such as robustness to model reduction errors when the model-based FTC architecture is implemented on the particulate process, as well as the problem of incomplete state measurements, are also discussed. Finally, the proposed methodology is applied to robustly stabilize a continuous crystallizer with a fines trap under control actuator faults.

2. Particulate processes with uncertainty, constraints and actuator faults

2.1 Process modeling using population, mass and energy balances
We focus on spatially homogeneous particulate processes modeled by the following general nonlinear system of partial integro-differential equations:

$$\frac{\partial n}{\partial t} = -\frac{\partial(G(x,r)n)}{\partial r} + w(n,x,r,\theta(t)) + g_1(n,x,r)\left[u_1^k + f_{a1}^k\right], n(0,t) = b(x(t))$$

$$\dot{x} = f(x) + g_2(x)\left[u_2^k + f_{a2}^k\right] + \chi\left(x,\theta(t), \int_0^{r_{max}} q(n,x,r)dr\right) \tag{1}$$

$$\left\|u_i^k(t)\right\| \le u_{i,max}^k, i = 1,2; \left\|\theta(t)\right\| \le \theta_b, k(t) \in \{1,2,...,L\}, L < \infty$$

where *n(r,t)* is the particle size distribution function, which is assumed to be continuous and sufficiently smooth in its arguments, $r \in [0,r_{max})$ is the particle size, *t* is the time, $x \in R^n$ is the vector of state variables that describe properties of the continuous phase (e.g., solute concentration, temperature and pH in a crystallizer), $u_1^k \in R$ is the manipulated input associated with the particulate phase (e.g., fines destruction rate), $u_2^k \in R$ is the manipulated input associated with the continuous phase (e.g., solute feed concentration in a crystallizer), $u_{i,max}^k$ is a real number that captures the size of constraints on the magnitude of the *i-th* manipulated input, $f_{a,i}^k \in R$ denotes a fault in the *i-th* control actuator, *k(t)* is a discrete variable that denotes which control configuration is active at time *t*, $\theta \in R^q$ denotes the vector of uncertain process variables (e.g., unknown process parameters and time-varying external disturbances), θ_b is a known bound on the size of the uncertainty. In the population balance of Eq.1, *G(x,r)* is the growth rate, $w(n,x,r,\theta) := w_0(n,x,r) + w_u(n,x,r)\theta$ is the net rate of introduction of new particles into the system, and $f(\cdot),q(\cdot),g_1(\cdot)g_2(\cdot),\chi(\cdot)$ are smooth nonlinear functions. The term containing the integral accounts for mass and heat transfer from the continuous phase to all the particles in the population and has the form $\chi(\cdot,\cdot,\theta) = \chi_0(\cdot,\cdot) + \chi_u(\cdot,\cdot)\theta$.

2.2 Model order reduction using the method of weighted residuals
To derive a finite-dimensional system that captures the dominant dynamics of the distributed parameter system of Eq.1, we first expand the solution in an infinite series in

terms of an orthogonal and complete set of basis functions, $\{\phi_k(r): r \in [0, r_{max})\}$, as $n(r,t) = \Sigma_{k=1}^{\infty} a_k(t)\phi_k(r)$, where $a_k(t)$ are time-varying coefficients. Substituting the series expansion into Eq.1, multiplying the population balance with appropriate weighting functions, integrating over the entire particle size spectrum, and finally truncating the series expansion up to order N (keeping the first N equations), we obtain the following approximate finite-dimensional system:

$$\dot{z}_i = h_i(z_1, z_2) + m_i(z_1, z_2)\left[u_i^k + f_{ai}^k\right] + d_i(z_1, z_2)\theta, i = 1,2 \qquad (2)$$

where $z_1 = a^N = [a_1^N \ ... \ a_N^N]^T$, $z_2 = x_N$, x_N and a_k^N are the approximations of x and a_k, respectively, obtained by an *N-th* order truncation of the infinite series, h_i, m_i, d_i are nonlinear functions whose explicit forms are omitted for brevity. The asymptotic validity of the approximation can be justified using standard results from perturbation theory (Christofides, 2002). In the next section, we present the design methodology for the FDI-FTC structure on the basis of the reduced-order system of Eq.2, and then discuss in Section 4 its implementation on the infinite-dimensional system of Eq.1.

3. Methodology for integrated robust FDI and fault-tolerant control

3.1 Bounded Lyapunov-based controller synthesis

The objectives of this step are to: (a) synthesize, for each control actuator configuration, a family of feedback controllers that enforce (in the absence of faults) constraint satisfaction and robust stability with an arbitrary degree of asymptotic attenuation of the effect of uncertainty on the closed-loop system, and (b) explicitly characterize the robust stability region associated with each configuration in terms of the constraints and the size of the uncertainty. To this end, we first design the controllers assuming the availability of accurate full-state measurements and then analyze the effects of measurement errors to derive sufficient conditions for closed-loop stability. Specifically, let V_i be a robust control Lyapunov function (Freeman and Kokotovic, 1996) for the *i-th* subsystem in Eq.2, and consider the feedback control law given by:

$$u_i^k(z) = -\left(\frac{\alpha_i(z) + \sqrt{\alpha_i^2(z) + \left(u_{i,max}^k \beta_i^k(z)\right)^4}}{\left(\beta_i^k(z)\right)^2\left[1 + \sqrt{1 + \left(u_{i,max}^k \beta_i^k(z)\right)^2}\right]}\right) L_{gi}V_i, i = 1.2, k(t) \in \{1,2,...,L\} \qquad (3)$$

where $\alpha_i = L_{hi}V_i + (\rho_i|z_i| + \chi_i\theta_b \ |L_{di}V_i|)(|z_i|/(|z_i| + \phi_i^k))$, $\beta_i^k = |L_{mi}V_i|$, $L_{hi}V_i$, $L_{mi}V_i$ and $L_{di}V_i$ are Lie derivatives of V_i, $z = [z_1 \ z_2]^T$ and $\rho_i > 0$, $\chi_i > 1$, $\phi_i^k > 0$ are adjustable parameters. It can be verified that each controller satisfies the control constraints within the state space region described by the set $\Pi_i^k := \{z \in R^{n+N}: \alpha_i (z, \ \rho_i \chi_i \ \phi_i^k, \ \theta_b) \le u_{i,max}^k \ \beta_i^k(z) \}$. To account for measurement errors, let $z_m(t) = z(t) + e(t)$ be the actual state measurement used to implement the controllers, where $e(t)$ is a bounded time-varying function $(|e(t)| \le \mu)$ that captures the measurement error. From the continuity of the control laws, it follows that the error in the implemented control action is also bounded, i.e., $|u_i^k(z_m) - u_i^k(z)| \le M_i(\mu)$ for some class $\mathcal{K}$ function M_i (a function is said to be of class $\mathcal{K}$ if it is non-decreasing and vanishes at zero; see Khalil (2002)). To ensure that the control constraints continue to be satisfied, evolution of the closed-loop states must be restricted within the sub-region defined by $\Lambda_i^k := \{z \in \Pi_i^k: |u_i^k(z)| \le u_{i,max}^k - M_i(\mu)\}$. Specifically, it can be shown using a Lyapunov argument that if $z(t) \in \Lambda_i^k$ and $f_{ai}^k(t) = 0$, for some t and i, then for sufficiently small ϕ_i (and/or sufficiently large χ_i), the time-derivative of V_i along the trajectories of the *i-th* fault-free closed-loop subsystem of Eqs.2-3 satisfies:

$$\dot{V}_i(z(t)) \leq -\gamma_i V_i(z_i(t)) + \sigma_i(\phi_i) + \Xi_i(\mu) := B_i(z_{m,i}(t), \gamma_i, \phi_i, \mu) \tag{4}$$

for some constant $\gamma_i > 0$ and some class $\mathcal{K}$ functions σ_i and Ξ_i. Furthermore, if $f^k_{a,i} = 0$ for all i and the closed-loop system is initialized within any invariant subset of $\Lambda^k := \Lambda^k_1 \cap \Lambda^k_2$, e.g., $z(0) \in \Omega^k_s(\theta_b, u^k_{max}) := \{z \in \Lambda^k : V_1(z) + V_2(z) \leq \delta_s\}$ for some $\delta_s > 0$, then each controller in the *k-th* control configuration (implemented with measurement errors) satisfies the control constraints and drives the trajectory of its corresponding subsystem in finite-time into a small neighborhood around the desired steady-state where it remains confined for all future times. The size of this terminal set can be made small by proper tuning of the controller parameters, but is ultimately constrained by the size of the measurement error μ. The set $\Omega^k_s(\theta_b, u^k_{max})$ therefore represents an estimate of the robust stability region starting from where practical closed-loop stability is guaranteed.

3.2 Robust fault detection and isolation scheme

Since the inequality of Eq.4 provides an explicit characterization of the expected fault-free behavior of each subsystem, it can be used to derive rules for FDI. Specifically, any breach of the given bound while $z(t) \in \Omega^k_s(\theta_b, u^k_{max})$ is an indicator of a fault in the corresponding actuator. However, since verifying this requires monitoring the state, which is known with only limited accuracy, we introduce two key modifications to safeguard against possible false alarms due to measurement errors. The modifications include (1) limiting the monitoring region to an appropriate subset of the stability region, i.e., $\Omega^k_c \subset \Omega^k_s$ where $z_m(t) \in \Omega^k_c \Rightarrow z(t) \in \Omega^k_s$ (this is needed to be able to infer the location of z in the state-space from the available measurement, z_m) and (2) enlarging the bounds on the dissipation of the Lyapunov functions. Specifically, by exploiting the continuity of both V_i and its time-derivative in their arguments, one can show that given any desired subset $\Omega^k_c \subset \Omega^k_s$, there always exists a class $\mathcal{K}$ function $X_i > \Xi_i$ such that:

$$\dot{V}_i(z_m(t)) \leq -\gamma_i V_i(z_{m,i}(t)) + \sigma_i(\phi_i) + X_i(\mu) := B^*_i(z_{m,i}(t), \gamma_i, \phi_i, \mu) \tag{5}$$

Note that any breach of Eq.5 while $z_m(t) \in \Omega^k_c$ implies that Eq.4 is breached while $z(t) \in \Omega^k_s$. Therefore, by using the time-derivative of V_i as a residual, and the right hand side of Eq.5 as an alarm threshold, a fault in the *i-th* actuator can be declared at time T_d if $z_m(T_d) \in \Omega^k_c$ and $\dot{V}_i(T_d) > B^*_i(z_{m,i}(T_d), \gamma_i, \phi_i, \mu)$. Comparing Eqs.4 and 5, wee see that the alarm threshold is essentially increased to ensure that any discrepancy between the actual and fault-free residuals exceeds what can be accounted for by measurement errors alone, and thus is due solely to faults.

3.3 Control system reconfiguration

Following FDI, the supervisor needs to determine based on the available (inaccurate) measurements which backup configuration can be activated to maintain robust closed-loop stability. To this end, consider the closed-loop system of Eqs.2-3 with $k(0)=j$, $z(0) \in \Omega^k_s$ and let T_d be the earliest time that a fault is detected in the *i-th* actuator. Then choosing the configuration $k(t)=v$ for $t \geq T_d$ where $z_m(T_d) \in \Omega^v_c(\theta_b, u^v_{max})$ ensures that the origin of the closed-loop system remains practically stable. Recall from the definition of Ω^k_c in Section 3.2 that $z_m(t) \in \Omega^k_c \Rightarrow z(t) \in \Omega^k_s$; therefore, the switching logic ensures that the fallback actuator configuration that is activated and implemented following FDI is one whose stability region contains the state at the time of switching. In the event that multiple configurations satisfy this stability condition, additional performance criteria can be introduced to aid in choosing the appropriate configuration.

4. Implementation issues

While the FDI-FTC structure presented above is designed on the basis of an approximate reduced-order model of the particulate process, it can be shown using regular perturbation techniques (e.g., see El-Farra and Giridhar (2008)) that it continues to enforce the desired stability and fault-tolerance properties in the infinite-dimensional system of Eq.1 provided that the reduced-order model is of a sufficiently high-order and the FDI rules are appropriately modified to account for the approximation errors. Specifically, the closeness of solutions between the approximate and infinite-dimensional systems can be exploited to obtain new alarm thresholds that can be chosen as close as desired to the bounds obtained for the approximate system for sufficiently large N. Furthermore, the problem when only measurements of the principal moments of the particle size distribution and the continuous-phase variables are available for on-line implementation can be addressed by incorporating a suitable nonlinear state observer to generate estimates of z from the measured outputs, which are then used to implement the controllers, FDI and reconfiguration laws. The observer must be designed to ensure sufficiently fast convergence of the state estimates. This property is necessary to ensure that the same diagnostic and fault-tolerance properties obtained under full-state feedback control can be practically recovered under output feedback control by proper choice of the observer tuning parameters (see El-Farra and Giridhar (2008) for an example observer design that satisfies this requirement).

5. Case study: Application to a continuous crystallizer with a fines trap

To demonstrate the implementation of the proposed FDI-FTC methodology, we consider in this section an example involving an isothermal continuous crystallizer with a fines trap. The trap is used to remove small crystals and increase the mean crystal size. The process model and parameters can be found in Christofides (2002). The control objective is to regulate the crystal and solute concentrations at desired set points by manipulating the flow rate of suspension through the fines trap and the inlet solute concentration in the presence of (1) actuator constraints and faults, (2) time-varying errors in the crystal and solute concentration measurements, and (3) time-varying uncertainties in the nucleation rate and the density of crystals. To ensure actuator fault-tolerance, a backup actuator is assumed to be available for each manipulated input for use in the event that the primary actuator fails. Using the method of moments and approximating the size distribution function using a Laguerre polynomial expansion to close the set of moment equations (Lei et al., 1971), an approximate fifth-order nonlinear ODE system describing the evolution of the first four moments of the crystal size distribution and the solute concentration is obtained and used for the synthesis of the FDI-FTC structure which is then implemented on a sufficiently high-order discretization of the process model obtained using a finite-difference scheme. The model derivation and controller synthesis details are omitted due to space limitations.

To detect and isolate faults during crystallizer operation, we consider the following two residual signals. The first is dedicated to faults in the first manipulated input (flow rate through fines trap) and captures the discrepancy between the actual and fault-free evolutions of z_1 in Eq.2 (the fault-free behavior is estimated from Eq.5). The second residual is dedicated to faults in the second manipulated input (solute feed concentration), and captures the discrepancy between the actual and fault-free evolutions of z_2 in Eq.2. To account for the effects of uncertainties, state estimation, measurement and model reduction errors in the absence of faults, the alarm thresholds for FDI are set at $\delta_1 = \delta_2 = 0.012$ by properly tuning the controllers and observers. To

demonstrate how the integrated FDI-FTC scheme works, the process is initialized using the healthy controllers which, as can be seen from the solid profiles in the top plots in Fig.1, successfully steer the crystal (left) and solute (right) concentrations near the desired set points very quickly. At *t=15 hr*, a fault is introduced in the actuator manipulating the suspension flow rate through the fines trap (valve stuck in the closed position; see dashed line in the left middle plot of Fig.1). The bottom plots in Fig.1 depict how this failure is detected and isolated since it causes the corresponding residual, r_1, to cross the specified threshold at *t=15.5 hr*, while practically not influencing r_2. Following FDI in the first controller, the supervisor switches to a backup actuator that maintains the closed-loop outputs near their desired set-points, as can be seen by the solid profiles in the top plots in Fig.1. After recovering from the first fault, a second failure in the actuator manipulating the solute feed concentration occurs at *t=25 hr* as shown by the dashed line in the right middle plot of Fig.1. By examining the residual profiles, it can be seen that this failure is detected and isolated in a timely manner since it causes r_2 to cross the threshold at *t=25.07 hr*, while r_1 remains within its threshold limit. Following the detection and isolation of the second actuator failure, the supervisor switches to the available backup actuator to preserve robust closed-loop stability. The dashed profiles in the top two plots of Fig.1 show the destabilizing effect of the two consecutive faults when no corrective action is taken.

Fig.1. Evolution of the crystallizer's controlled outputs (top), manipulated inputs (middle) and FDI residuals (bottom) under two consecutive failures in the control actuators and subsequent reconfiguration.

References

P.D. Christofides, 2002, *Model-Based Control of Particulate Processes*, Netherlands: Kluwer Academic Publishers.

N.H. El-Farra and A. Giridhar, 2008, Detection and management of actuator faults in controlled particulate processes using population balance models, *Chem. Eng. Sci.*, 63, 1185-1204.

R.A. Freeman and P.V. Kokotovic, 1996, *Robust Nonlinear Control Design: State-Space and Lyapunov Techniques*, Boston: Birkhauser.

H.K. Khalil, 2002, *Nonlinear Systems*, Upper Saddle River, NJ: Prentice Hall.

P. Larsen, D. Patience, and J.B. Rawlings, 2006, Industrial crystallization process control, *IEEE Contr. Syst. Mag.*, 26, 70-80.

S.J. Lei, R. Shinnar, and S. Katz, 1971, The stability and dynamic behavior of a continuous crystallizer with a fines trap, *AIChE J.*, 17, 1459-1470.

M.J. Park, M.T. Dokucu, and F.J. Doyle, 2004, Regulation of the emulsion particle size distribution to an optimal trajectory using partial least squares model-based predictive control, *Ind. Eng. Chem. Res.*, 43, 7227-7237.

D. Semino and W.H. Ray, 1995, Control of systems described by population balance equations-II: emulsion polymerization with constrained control action, *Chem. Eng. Sci.*, 50, 1825-1839.

G. Zhang and S. Rohani, 2003, On-line optimal control of a seeded batch cooling crystallizer, *Chem. Eng. Sci.*, 58, 1887-1896.

1359

Iterative Feedback Tuning of State Space Control Loops with Observers Given Model Uncertainty.

Jakob K. Huusom,[a] Niels K. Poulsen,[b] Sten Bay Jørgensen[a]

[a]*CAPEC, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 229, DK-2800 Lyngby, Denmark, jkh@kt.dtu.dk;sbj@kt.dtu.dk*
[b]*Department of Informatics and Mathmatical Modelling, Technical University of Denmark, Buliding 321, DK-2800 Lyngby, Denmark, nkp@imm.dtu.dk*

Abstract

Direct tuning is investigated as an alternative to iterative model estimation and control design for state space systems in case of unsatisfactory loop performance. Direct tuning of the model parameters in the feedback control and the observer design by Iterative Feedback Tuning, optimize loop performance when the initial designs are based on an uncertain model estimate. The certainty equivalence design is kept throughout the iterations when the model parameters is updated. This methodology constitutes a promising algorithm when developing tools for online tuning of state space control systems.

Keywords: LQG control, Iterative Feedback Tuning.

1. Introduction

The need for optimal process operation has rendered methods for optimization of control loop parameters an active research area. Much attention has been directed on performing control oriented system identification, which implies model estimation from closed loop data [1,2]. Optimizing the parameters in a control loop is an iterative procedure since the data from one experiment will depend on the current controller, and repeated iteration is necessary for the loop performance to converge to a minimum. An alternative would be a direct data driven approach to tuning without utilizing a model estimate. Data driven tuning methods have mainly been reported for systems given in transfer function form. The most established method is Iterative Feedback Tuning [3]. This method optimizes the closed loop performance by adjusting the control parameters through a gradient based scheme. The gradient of the cost function is replaced by an unbiased estimate evaluated from special closed loop experiments. Direct tuning is often computationally less demanding than identification and model based control design. Direct tuning methods can be used when insufficient knowledge of the model structure limits nominal performance, where the system is tuned based on the certainty equivalence principle.

This paper investigates the use of the direct tuning method, Iterative Feedback Tuning, for optimization of the parameters in the feedback loop and the state observer for a control loop based on a state space system description. Based on the certainty equivalence principle, analytical solutions for optimal values of the feedback and Kalman gain exist. This assumption renders the loop performance sensitive to model

errors and bias. Direct controller tuning of the uncertain model parameters may serve as an interesting alternative, when fine tuning a control loop is desired or when degrading loop performance is observed.

This paper is organized as follows. In the following section a short introduction to the system and control loop description is given. Section 3 discusses optimal control and tuning. In Section 4 an illustrative simulation example is given before final conclusions eventually are drawn.

2. The state space control loop

Given the following linear, discrete time, single input/single output, time-invariant system description:

$$x_{t+1} = Ax_t + Bu_t + e_t^P, \qquad e_t^p \in N(0, P_{e^P})$$
$$y_t = Cx_t + e_t^m, \qquad\qquad e_t^m \in N(0, \sigma_{e^m}^2) \tag{1}$$

Where x_t represents the system states, u_t is the manipulated variable and y_t is the measurement at discrete time instants. e_t^p represents process noise and e_t^m is measurement noise. Both e_t^p and e_t^m will be assumed white and the cross correlation between these types of noise will be assumed zero. It is desired to control this system using the feedback law:

$$u_t = -Lx_t + Mr_t \tag{2}$$

Where L is a constant feedback gain matrix and M is a controller gain for the reference signal. Since the exact values of the states are not known, an observer is used to generate the state estimates used in the control law based on measurements of the process output and a process model. The observer has the form of the predictive Kalman filter with the constant gain matrix K, assuming stationary conditions.

$$\hat{x}_{t+1|t} = \hat{A}\hat{x}_{t|t-1} + \hat{B}u_t + K(y_t - \hat{C}\hat{x}_{t|t-1}) \tag{3}$$

The structure of the state space feedback loop consisting of equation (1) and (3) is shown in Figure 1.
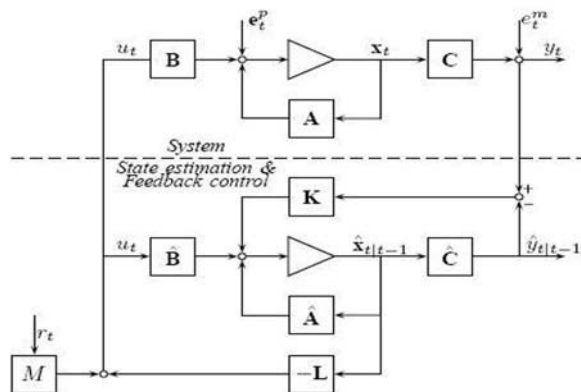
Figure 1. Block diagram of the closed loop state space system with state observer

In order to ensure a static gain from the reference to the estimated output of unity, the following requirements can be derived:

$$M = [\hat{C}(I - \hat{A} + \hat{B}L)^{-1}\hat{B}]^{-1} \qquad (4)$$

Introducing the state estimation error: $\tilde{x}_t = x_t - \hat{x}_t$, provides a convenient description with a more clear distinction between feedback control and state estimation dynamics [4]. If the system (1) is stabilizable and detectable a set $\{L,K\}$ exists which renders the system stable [5]. Hence if optimal values for the feedback and Kalman filter gains are used stability is guaranteed.

3. Optimal control and tuning

Optimal values for both the observer gain K and the feedback gain L exist and have known analytical solutions if the true system is known [6]. The optimal, stationary value for the gain matrix in the Kalman filter can be evaluated based on the process model and information of the noise intensities, from an algebraic Riccati equation. The optimal controller gain will also depend on the optimization criterion. In this paper the control design will minimize the value of a cost function for the loop performance for a single input/single output system:

$$F(y,u) = \frac{1}{2N}\sum_{t=1}^{N} y_t^2 + \lambda u_t^2 \qquad (5)$$

Where λ determines the weighting between the penalty on the output and the control. For optimal tracking the output is replaced by the tracking error in the cost function. The optimal Linear Quadratic Gaussian controller (LQG) produces an optimal feedback gain which minimizes the quadratic cost function:

$$F_{LQG}(y,u) = \frac{1}{2N}\sum_{t=1}^{N} \hat{x}_t^T Q_R \hat{x}_t + \lambda u_t^2 \qquad (6)$$

using the linear system description in (3) and assuming that N approaches infinity. As for the observer gain, the solution is given by an algebraic Riccati equation. In case $Q_R = C^T C$ the cost function (6) is equivalent to (5).

In absence of an accurate process model, direct tuning of the closed loop can provide the optimal control parameters. It has been shown in [7] that Iterative Feedback Tuning can be applied to a state space control system with state observer. The iterative procedure converges towards the optimal values for the two gains, when the observer is constructed from full model information. This result has more academic than practical interest since analytical solutions are known when a model is available. It is on the other hand interesting to investigate the potential for using direct tuning in case of model uncertainty. Parametric model uncertainty will give the following equations for the state and the estimation error:

$$x_{t+1} = (A - BL)x_t + BL\tilde{x}_t + BMr_t + e_t^P$$

$$\tilde{x}_{t+1} = (\hat{A} - K\hat{C})\tilde{x}_t - Ke_t^m + e_t^P + \Delta \qquad (7)$$

$$\Delta = [(A - \hat{A}) - K(C - \hat{C})]x_t + (B - \hat{B})[Mr_t - L(x_t - \tilde{x}_t)]$$

Only if the true process and the model estimate are equivalent will the term Δ be zero. In this case the optimal closed loop performance can be achieved by solving the Riccati equations for the two gains. In case of parametric uncertainty, a certainty equivalent control design will not produce optimal closed loop performance. Hence either model re-estimation with a subsequent update of the feedback and observer gain or direct tuning can be used for optimization. In [8] direct tuning of the two gains was performed using Iterative Feedback Tuning in order to investigate to which extent adjusting the gains could compensate for the erroneous model parameters in the state estimator. The loop performance was improved iteratively, and in some cases the performance degradation caused by an erroneous parameter could be completely compensated. This paper will move further and tune the model parameters in the state estimator by direct tuning. Given an updated set of model parameters, new optimal feedback and Kalman filter gains are recalculated from the Riccati equations.

4. Case study

In order to illustrate the potential of using direct tuning in form of the Iterative Feedback Tuning method the discrete time, scalar, state space system with state observer (8) is investigated. This system is too simple to have any industrial relevance but the objective of this example is to show the principles in the tuning method and it's ability to converge to optimality. The general method described in the paper concerns higher order linear models too.

$$x_{t+1} = 0.98x_t + 0.02u_t + e_t^P, \qquad e_t^P \in N(0,1^2)$$

$$y_t = 1x_t + e_t^m, \qquad\qquad e_t^m \in N(0,0.01^2) \qquad (8)$$

In the optimization criterion $\lambda = 0.001$ is used. In the case when only the noise variance is unknown and all other parameters in the observer are correct, only the Kalman filter gain needs to be tuned. In case any of the parameters a, b or c in model estimate are erroneous, it will affect both the values of the Feedback and Kalman gains and the state estimate. In the following, three experiments are performed where one of the parameters a, b or c in the state estimate is wrong respectively. By direct tuning of the parameter in question and subsequent update of the Feedback and Kalman gains, to ensure optimality by certainty equivalence in each iteration, the closed loop performance is optimized. Fig. 2, 3 and 4 show the results from 15 iterations of the tuning, when the erroneous parameters is either too large or too small. All experiments are able to converge to the optimal solution in 5-10 iteration which is very good when tuning is conducted for the disturbance rejection case. The method also allows tuning of all the model parameters simultaneously and for multivariable systems, but these results have been omitted due to the spatial constraints of this contribution. It is however ongoing work.
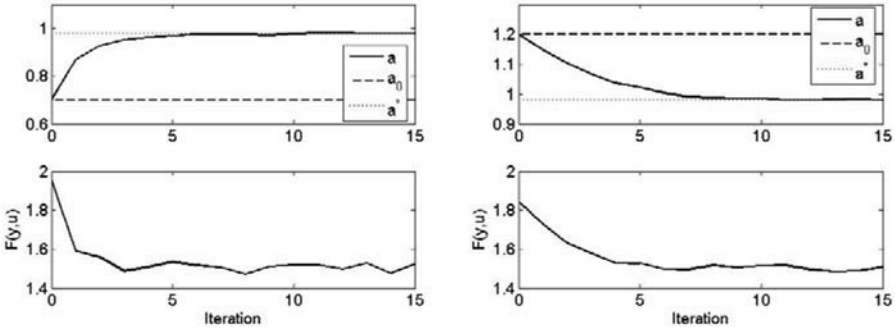
Figure 2. 15 iterations by the tuning method when the parameter *a* in the observer is erroneous. Both the iteration of the parameter value and the corresponding closed loop performance are given.
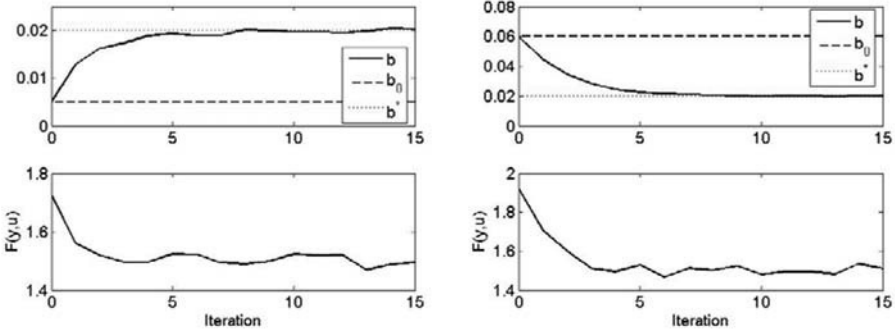
Figure 3. 15 iterations by the tuning method when the parameter *b* in the observer is erroneous. Both the iteration of the parameter value and the corresponding closed loop performance are given.
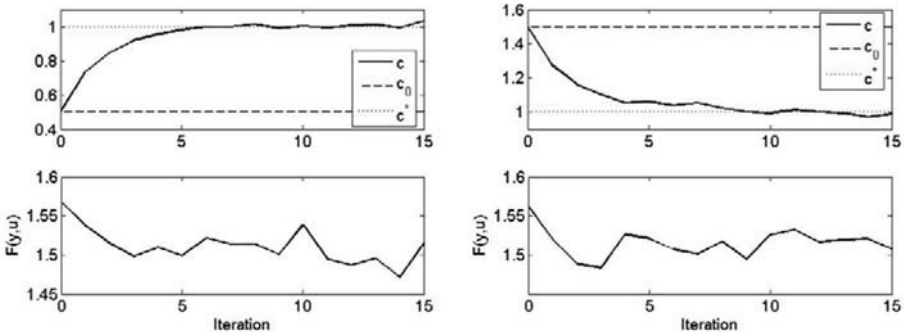
Figure 4. 15 iterations by the tuning method when the parameter *c* in the observer is erroneous. Both the iteration of the parameter value and the corresponding closed loop performance are given.

The cost function gradient estimate used in a Newton scheme by the Iterative Feedback Tuning method, utilizes a transfer function description of the state space control loop on Fig. 1 which is given in [7]. This estimate is constructed by filtering closed loop

input/output data through a filter which contains the gradient of the controller with respect to the tuning parameters. Since the feedback and the feed forward controller in the transfer function description of the state space control loop are functions of both the model estimate and the gain matrices, the partial derivatives of the optimal gains are needed with respect to the tuning parameters. These have been obtained by a first order forward difference approximation in the results presented here.

The results presented are by far superior to the results in [8] where gains were tuned while the erroneous parameters in the state estimator were kept constant. The increased complexity of tuning the system parameters and adapting the gains seems to be rewarded since convergence of the model estimate to the true system leads to optimality for the closed loop performance.

5. Conclusions

Direct tuning in form of Iterative Feedback Tuning has been used to adjust the system model parameters in the observer in a state space control loop in order to minimize the closed loop performance cost. When the model in the control and the observer design is iteratively updated so are the Feedback and Kalman gains in order to ensure optimality by certainty equivalence in the design. The tuning parameters converged to the true system hence the achieved closed loop performance converged to optimality.

Acknowledgements

References

1. R. J. P. Schrama, Accurate identification for control: The necessity of an iterative scheme. IEEE Transactions on automatic control, vol. 37, (7), pp. 991-994, 1992.
2. M. Gevers, A decade of progress in iterative process control design: from theory to practice, Journal of process control, vol. 12,(4), pp. 519-531, 2002.
3. H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, Iterative feedback tuning: Theory and applications, IEEE Control Systems Magazine, vol. 18, (4), pp. 26-41, 1998.
4. K. J. Åström, Introduction to Stochastic Control Theory, Academic Press, 1970.
5. H. Kwakernaak and R. Sivan, Linear Optimal Control Systems. John Wiley & Sons,1972.
6. B. D. O. Anderson and J. B. Moore, Optimal Control. Linear Quadratic Methods, Prentice-Hall Int. Ed., 1989.
7. J. K. Huusom, N. K. Poulsen and S. B. Jørgensen, Data Driven Tuning of State Space Controllers with Observers. Submited for ECC 2009
8. J. K. Huusom, N. K. Poulsen and S. B. Jørgensen, Data Driven Tuning of State Space Control loops with unknown state information and model uncertainty. Accepted for ESCAPE19 2009.

1365

Real Time Optimization (RTO) with Model Predictive Control (MPC)

Glauce De Souza[a], Darci Odloak[b], Antônio C. Zanin[c]

[a]Department of Chemical & Materials Engineering, The University of Auckland, 20, Symonds Street 8th floor, City Campus 1042, Auckland, New Zealand
[b]LSCP/CESQ - Department of Chemical Engineering - University of São Paulo, Av. Prof. Luciano Gualberto, n. 380, trav. 3, CEP 05508-900, São Paulo - SP, Brazil. Brazil.
[c]PETROBRAS, Brazilian Oil Company

Abstract:

This paper studies a simplified methodology to integrate the real time optimization of a continuous system into the model predictive controller in the one layer strategy. The gradient of the economic objective function is included in the cost function of the controller. One of the control objectives is to zero the reduced gradient of the economic objective while maintaining the system outputs inside their zones. Optimal conditions of the process at steady state are searched through the use of a rigorous nonlinear process model, while the trajectory to be followed is predicted with the use of a linear dynamic model that can be obtained through a plant step test. Moreover, the reduced gradient of the economic objective is computed taking advantage of the predicted input and output trajectories. The main advantage of the proposed strategy is that the resulting control/optimization problem can be solved with a quadratic programming routine at each sampling step. Simulation results show that the approach proposed here is comparable to the strategy that solves the full economic optimization problem inside the MPC controller where the resulting control problem becomes a nonlinear programming with a high computer load.

Keywords: Real Time Optimization, Model Predictive Control, Fluid Catalytic Converter, Non-linear programming

1. Introduction

Typically, the RTO application optimizes the process operating conditions and updates the set-points to local MPCs, which are based on linear dynamic models. A steady-state RTO may not be sufficient if, for instance, the set of active constraints with significant economic importance changes frequently. For this type of process, it is more suitable to use dynamic optimization with a non-linear model, which may be achieved using dynamic RTO (DRTO) or non-linear MPC with economic objective (Kadam et al., 2003; Tosukhowong et al., 2004). However, the dynamic optimization of a complex process is still not achievable with the available computing resources (Lu, 2003). Aske et al. (2008) propose the use of a coordinator MPC, which explores the degrees of freedom of the system in order to maximize throughputs at the bottlenecks of the system. Economic optimization studies for

oil refining and chemicals production systems have proven to be very beneficial (Georgiou et al., 1997: Rotava & Zanin, 2005). Similar algorithms to the ones of the petrochemical industry were applied by Mercangoz & Doyle III (2007) to the pulp and paper manufacturing system. To maximize the economic benefit while maintaining the desired product quality may be hard in a continuous production environment with frequent changes (Engell, 2007). Here, the focus is on the integration of real time optimization and control of continuous processes through the strategy called one-layer approach (Zanin et al. 2002) or full optimizing control (Rolandi & Romagnoli, 2005).

In process plants, the economically optimal operation is usually addressed by a two-layer structure (Marlin & Hrymak, 1996). In the upper layer, the operating point or steady state of the plant is optimized based upon a rigorous nonlinear stationary plant model. This two layer structure has some drawbacks, as optimization is only performed intermittently at a low sampling rate and inconsistencies may arise from the use of different process models at the two layers. These issues are partly addressed by schemes in which the economic optimization is integrated within a linear MPC controller. Typical sampling times for the RTO layer in a two layer approach are of the order of several hours (Engel, 2007) while the control layer provides tracking and disturbance rejection on shorter time-scales from seconds to minutes.

Zanin et al. (2002) included the economic objective computed at the predicted steady state as an additional term of the cost function of the controller. Here it is presented a simplified strategy to integrate the RTO problem into the linear model predictive controller in the one layer approach. An approximation of the gradient of the economic function is included in the MPC cost function with an appropriate weight. Simulation results obtained with the nonlinear model of Moro & Odloak (1995) are used to compare the proposed approach to the approach of Zanin et al. (2002), and show that the results of the two approaches may be comparable.

2. The one-layer RTO/MPC controller

Zanin et al. (2002) propose to integrate the real time optimization of continuous processes into the model predictive controller by solving at each sampling step k the optimization problem:

$$\min_{u(k+i);\, i=0,...,m-1} \sum_{j=1}^{p} \left\| W_1 (y(k+j)-r) \right\|_2^2 + \sum_{j=0}^{m-1} \left\| W_2 \Delta u(k+j) \right\|_2^2 + W_3 f_{eco}\left(u(k+m), y(k+\infty)\right) \quad (1)$$

subject to

$$h\left(u(k+m), y(k+\infty)\right) = 0 \qquad (2)$$

$$u_{\min} \leq u(k+j) \leq u_{\max} \quad j=0,1,...,m-1 \qquad (3)$$

$$\Delta u_{\max} \leq \Delta u(k+j) \leq \Delta u_{\max} \quad j=0,1,...,m-1 \qquad (4)$$

$$y(k+j) = g_j\left(x(k), \Delta u_k\right), \quad j=1,...,p \qquad (5)$$

where

p and m are the output and input horizons, respectively; $u(k+j)$ is the control input computed at time k to be applied at time step $k+j$; $y(k+j)$ is the output prediction at time step $k+j$; r is the desired value of the output; $\Delta u(k+j) = u(k+j) - u(k+j-1)$; W_1, W_2 and

W_3 are positive weights; f_{eco} is the economic objective corresponding to the predicted steady state of the process; $y(k + \infty)$ is the process output steady state computed with Eq. (2). Eq. (5) corresponds to the output prediction performed with the linear dynamic model of the system, and $\Delta u = \begin{bmatrix} \Delta u(k) & \Delta u(k+1) & \cdots & \Delta u(k+m-1) \end{bmatrix}^T$ is the control sequence.

In the RTO/MPC algorithm defined above, the outputs are controlled by zones, which means that the desired values r of the controlled outputs do not have fixed values.

Since in the general case, f_{eco} and h are nonlinear functions of the inputs and outputs of the system, the problem solved by the controller is a NLP while the problem solved in the conventional MPC is a QP. Although, the recent progress in the algorithms for the numerical solution of NLP problems, one can expect some difficulties in solving the RTO/MPC problem for large dimension systems. If the solver of the one layer RTO/MPC problem fails to produce a converged solution within the sampling period of the controller, no control action will be available to be implemented and the system will remain in open loop, which is highly undesirable. Thus, there is interest in developing an alternative solution to the one layer RTO/MPC problem that preserves some of the advantages of the one layer approach, but without delaying the computation of the optimizing control action.

3. The simplified one-layer RTO/MPC

Consider a multivariate system with *ny* outputs and *nu* inputs. At any time step *k*, suppose that the stationary prediction of the controlled output related to *u* is $\hat{y}$, and the economic function associated with this predicted steady state can be represented as follows:

$$F = f_{eco}(u, \hat{y}) \tag{6}$$

Assuming that the vector that represents the control action is changed to $u + \Delta \bar{u}$, the first order approximation to the gradient of the economic function can be represented as follows:

$$\varsigma_{u+\Delta u} = \left[\frac{\partial F}{\partial \hat{y}} K_p + \frac{\partial F}{\partial u} \right] + \left[K_p{}^T \frac{\partial^2 F}{(\partial \hat{y})^2} K_p + \frac{\partial^2 F}{\partial u \partial \hat{y}} K_p + \frac{\partial F}{\partial \hat{y}} \left(\frac{\partial^2 \hat{y}}{\partial u^2} \right) + K_p \frac{\partial^2 F}{\partial \hat{y} \partial u} + \frac{\partial F}{\partial u^2} \right] \Delta \bar{u}$$

where $K_p = \partial \hat{y} / \partial u$

The above equation can be represented as follows:

$$\varsigma_{u+\Delta u} = d + G \Delta \bar{u} \tag{7}$$

where $\Delta \bar{u} = u(k+m-1) - u(k-1)$ is the total move of the input vector.

Then, the simplified real time optimization integrated into the linear MPC controller (SRTO/MPC) proposed here has the following control objective:

$$\min_{u(k+i); \, i=0,\ldots,m-1} J = \sum_{j=1}^{p} \left\| W_1(y(k+j) - r) \right\|_2^2 + \sum_{j=0}^{m-1} \left\| W_2 \Delta u(k+j) \right\|_2^2 + \left\| W_3 \varsigma_{u+\Delta u}^T \right\|_2^2 \tag{8}$$

subject to: (3) to (5) and (7)

The integration of the economic function into the controller can push one or more outputs towards their bounds. To consider this effect in the calculation of the gradient of the economic objective, the following algorithm is followed:

At each sampling time k, vector d and matrix G are computed using the predicted steady state corresponding to the control input implemented at time step k-1. Then, the problem defined in (8) is initially solved assuming that no controlled variables will be constrained. Next, the resulting new predicted values for the controlled variables are checked in order to confirm that there are no violations of the constraints. If the prediction values of the outputs do not violate any constraint, the solution obtained through the solution to the problem represented in (8) is implemented in the real system. If the prediction of one of the outputs is such that $\hat{y}_i \geq y_{\text{max},i}$ or $\hat{y}_i \leq y_{\text{min},i}$. Then, the problem defined in (8) is solved again with the following additional constraint.

$$\Delta y_i = \sum_{j=1}^{nu} K_{p_{i,j}} \Delta \overline{u}_j = 0 \qquad (9)$$

The output predictions corresponding to the solution defined in the previous step should be checked to verify if any other output will break its bounds. If this happens additional constraints of the type represented in (9) should be included in the control problem. This procedure should be repeated until no other constraints are required to be included in the control problem.

It can be shown that the control cost function of the MPC with economic objective represented in (8) is a quadratic function of the control move. As the constraints represented in (3) to (5) and (7) are linear in the control, the optimization problem that produces the proposed MPC with economic objective is a QP, with a computer effort similar to the computer effort of the conventional MPC.

4. The FCC (Fluid Catalytic Cracking) unit

To compare the performance of the SRTO/MPC algorithm proposed in section 3 with the RTO/MPC of Zanin et al. (2002) summarized in section 2, it is considered here the same process system studied by Zanin et al. (2002). The system is the Fluid Catalytic Cracking Kellogg Model F, which is described in Moro and Odloak (1995).

In the simulations presented here, only a few variables of the industrial control system are considered. The main manipulated inputs are: u_1 (*ton/h*), the total air flow rate to the two-stage regenerator, u_2 (*%*), the valve opening of the regenerated catalyst and u_3 (*m³/h*), the gasoil feed flow. Considering the approach of zone control the controlled outputs considered here are: y_1 (*C*), the temperature in the dilute phase of the regenerator and y_4 (*C*), the riser temperature. Here, it is assumed that the economic objective of the FCC unit is to maximize the production of liquefied petroleum gas (LPG), which depends on the feed properties and process operating conditions. In the simulations the tuning parameters adopted in the RTO/MPC of section 2 are the following: $m=2$, $p=70$, $W_1=\text{diag}(0.2\ 0.1\ 0.1\ 1)$, $W_2=\text{diag}(5\ 5\ 5\ 5)$ and $W_3=-1$. The sampling period is equal to 1 min.

The proposed algorithm SRTO/MPC has the same tuning parameters and constraints as the RTO/MPC except the weight of the economic component, which in this case is $W_3= 0.3$.

Fig. 1 shows the behavior of the LPG production in m^3/h, which is the economic objective for the two approaches when the controller is started from the steady state. It is also shown the controlled outputs and the manipulated inputs. One may observe that the simplified algorithm (SRTO/MPC) captures most of the economic benefit obtained in the one-layer (RTO/MPC) algorithm proposed by Zanin et al. (2002). The increase in the economic term

is faster in the RTO/MPC, but the final value of the economic benefit is not significantly different in the two approaches. Obviously, in both methods the speed of response of the economic benefit will depend on the tuning weight W_3. The larger the absolute value of this parameter, the faster the economic benefit will respond to a change in the operating conditions or disturbances. However, as the constraints on the system outputs are not explicitly included in the control/optimization problem, the result of a large economic weight W_3 may be undesirable. If the optimal unconstrained operating point of the controlled system lies outside the convex set defined by the constraints, the economic term in the control cost function tends to push the system outputs to outside the control zones. In the example simulated here, Fig. 1 shows that y_1 tends to stabilize slightly below the minimum bound, while y_4 tends to stabilize slightly above the maximum bound. Also, input u_3 is pushed by the controller to its maximum bound. This indicates that all the degrees of freedom of the system have been consumed by satisfying the system constraints and the optimal operating point lies at the intersection of the constraints. A straightforward way to reduce the distance from the point, in which the outputs stabilize to the control zones, is by increasing the corresponding elements of weight W_1. However, as the dynamic model used in the controller is a linear model, there is a limit on the maximum value of this weight before the closed loop system becomes unstable. For instance, keeping the remaining tuning parameters of SRTO/MPC as defined above, the system becomes unstable if the component of W_1 corresponding to y_1 is made equal to 1. Thus, the integration of real time optimization with linear MPC in a one layer strategy seems to pose a robustness problem related to the uncertainty of the linear model.

5. Conclusion

Here, it is proposed a simplified version of the one layer RTO/MPC controller presented in previous works that integrates the two functions, real time optimization and multivariable constrained control, in a single layer. The control objective is to zero the reduced gradient of the economic objective while maintaining the system outputs inside their zones. The main advantage of the proposed approach is that the integrated control/optimization problem becomes a QP, instead of a NLP as in the previous approach. Simulation results have shown that a substantial percentage of the possible economic benefit associated with real time optimization can be captured with the proposed approach.

References

Aske, E. M. B., Strand, S., & Skogestad S. (2008). Coordinator MPC for maximizing plant throughput. *Computers and Chemical Engineering*, 32 (1-2), pp.195-204.

Engell, S. (2007), Feedback control for optimal process operation, *Jour. Proc. Control*, 17, 203-219.

Georgiou, A., Taylor, P., Galloway, R., Casey, L., Sapre, A. (1997). Plantwide closed loop real time optimization and advanced control of ethylene plant (CLRTO) improves plant profitability and operability. In *Proceedings of NPRA Comp. Conf., Nat. Petr. Ass.*, CC 97-139, New Orleans..

Kadam, J., Marquardt, W., Schlegel, M., Backx, T., Bosgra, O., Brouwer, P. J. (2003), Towards integrated dynamic real-time optimization and control of industrial processes. In *Proceedings Foundations of Computer-Aided Process Operations (FOCAPO2003)*, pp. 593-596.

Lu, J. (2003). Challenging control problems and emerging technologies in enterprise optimization. *Control Engineering Practice, 11,* 847-858.

Marlin, T.E., Hrymak, A. N. (1996). Real-time operations optimization of continuous processes. *In Proceedings of CPC V, AIChE Symposium Series*, vol. 93, 1997, pp.156-164.

Mercangoz, M., Doyle III, F. J. (2008). Real-time optimization of the pulp mil benchmark problem. *Computers and Chemical Engineering*, 32 (4-5), pp. 789-804.

Moro, L.F.L., Odloak, D. (1995). Constrained multivariable control of fluid catalytic cracking converters. *Journal of Process Control*, 5, 1-11.

Rolandi, P. A., Romagnoli, J. A. (2005). An integrated environment for support of process operations, *AIChE Annual Meeting, Conference Proceedings*, pp. 6026-6027.

Rotava, O., Zanin, A. C. (2005). Multivariable control and real-time optimization-An industrial practical view. *Hydrocarbon Processing, 84*(6), 61-71.

Tosukhowong, T., Lee, J., Lee, J., Lu, J. (2004). An introduction to a dynamic plant-wide optimization strategy for an integrated plant. *Computers and Chemical Engineering*, 29, 199-208.

Zanin, A. C., Tvrzská de Gouvêa, M., Odloak, D. (2002). Integrating real-time optimization into the model predictive controller of the FCC system. *Contr. Engng. Pract., 10*, 919-831.
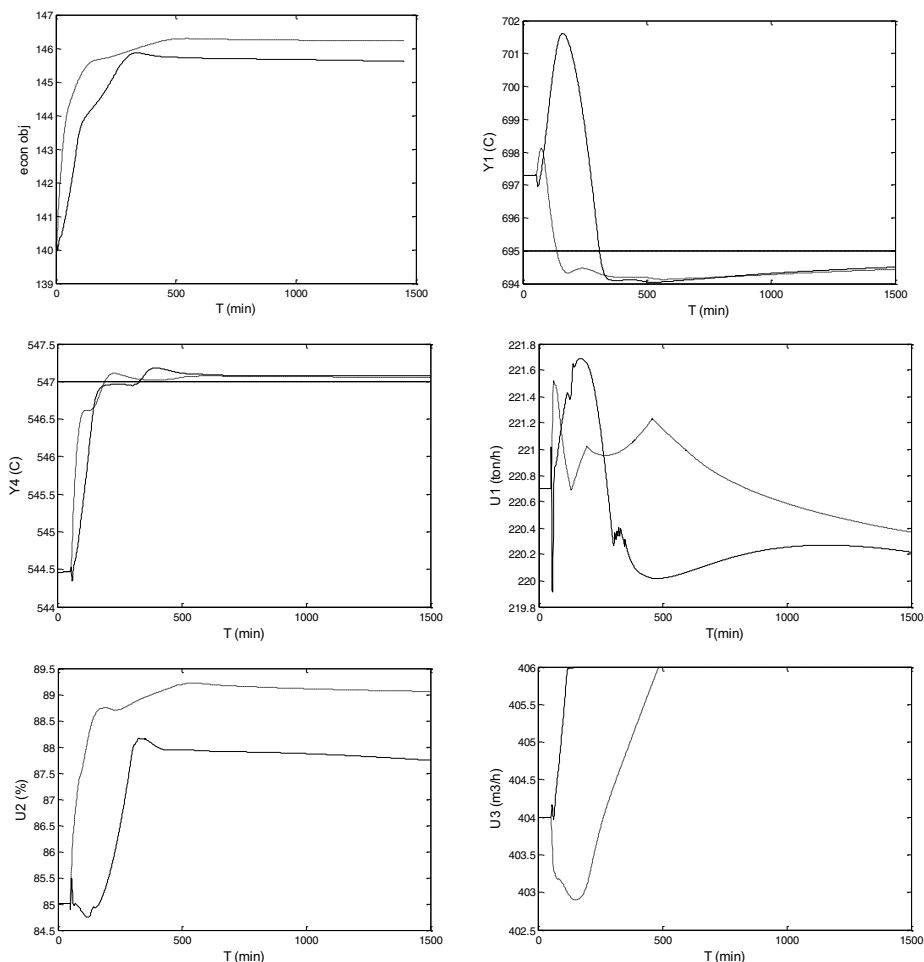
Fig. 1 FCC variables of SRTO/MPC (———) and RTO/MPC(− − −)

1371

Towards a Viscosity and Density Correlation for Dairy Fluids - A Soft Sensor Approach

Tien-I Lin[a], Glauce De Souza[a] and Brent Young[a*]

[a]*Chemical and Materials Engineering Department, The University of Auckland*
20 Symonds Street, Private Bag 92019, Auckland 1142, New Zealand
*Author to whom correspondence should be addressed: b.young@auckland.ac.nz

Abstract

In spray drying, the viscosity of the feed is a critical characteristic. It affects the droplet size formed at the atomiser, the end product properties (e.g. bulk density, moisture content and solubility) and the characteristics of the process (e.g. drying time and wall sticking behaviour.). Currently viscosity has been robustly measured online on a pilot scale plant but no robust online measurements were reported for industrial scale plant. However, it is possible that empirical correlations can be developed.

For a spray drying plant, the density is usually measured online and used to calculate the percentage of total solids. It has also been identified that there is a positive relationship between the viscosity of the feed concentrate and the density. However this correlation is not reported in the open literature. Therefore, it is of interest to attempt to identify and validate this correlation for model predictive control purposes. Due to the availability and reliability of on line density measurement at the industrial plant, the validation and application such a correlation is potentially relatively simple.

A pilot plant was set up with a Micro-Motion density meter and a Hydromotion vibrational online viscometer. The viscometer was placed in a water- jacketed tank, where the measurements took place. Milk concentrate for total solids (TS) contents between 35% to 45% was then re-circulated in the system at temperatures up to 70°C at atmospheric pressure to obtain online measurements of both density and viscosity.

Keywords: *Viscosity, Milk powder, Process control, Soft sensor*.

1. Introduction

Milk powder is the most important processed dairy product for export due to its long shelf life and ease of transportation. However, there is a huge energy requirement in the milk power production process, predominately from evaporation and drying

Spray drying is the most common milk powder manufacturing method. In recent years, spray drying technology has been developing in two separate directions. The improvement of energy efficiency and the development of new functional products by applying spray drying techniques. However the future success of both directions lies with better understanding and controllability of the process [1].

In the 2004 International Symposium on Spraying of Milk Products, advanced process control (APC) was considered to be one of the top areas that required further development due to difficulty in operating larger spray drying units. This area is currently advancing in a significant rate [1]. However, the primary challenge in developing a model for model predictive control remains the of understanding of

properties and stickiness behaviour of the droplet and powder in the spray drying process [1].

Currently, many spray drying process plants adopt process control schemes that are very dependent on the operator's experience, where the set point of a process is determined by the operators [2, 3]. Some attempts have been made to incorporate advanced process control but fail to take physical attributes of the milk into the process model [4], [5].

For spray drying, temperature is normally measured in lieu of moisture content online [4, 6, 7]. This method assumes a correlation between the temperature measurement and the evaporation rate [6]. However there is a fundamental flaw in this method. The variation in inlet air, wet bulb temperature or equilibrium constant of inlet air can result in dramatic change in moisture content of the product but have no effect on the outlet air temperature [6]. This relationship changes depending on the physical attributes of the milk. Chen pointed out that viscosity of the milk concentrate is important in this process [3].

Before any process control is applied to a system, a process model must be developed. For spray drying the main challenge is developing a process model for the spray drying unit. The relationship between the process condition and the product characteristics must be known. Also thermodynamic/fluid dynamic equations must be described in terms of online heat and mass transfer measurements [8]. These measurements must be valid for a range of process conditions and product specifications. Given these requirements, viscosity must be measured in terms of total solids and shear rate for a range of the product specifications to give accurate fluid dynamic equations.

In many cases, development of process control is limited by the availability of robust and accurate sensors. This can be due to the cost effectiveness and the limitation of technology. In the case of viscosity in the diary industry, attempts of using online viscometers and soft sensor approaches have both been undertaken [8, 9] but only on pilot plant. Density of milk concentrate is normally monitored at the end of the evaporator and used to determine the total solid content of the milk concentrate. It has been reported by several sources that a correlation between density and viscosity of milk concentrate exits, however without details.

This paper intends to describe an experimental set up that is capable of finding such a correlation. The performance of the experimental rig was evaluated and preliminary results are also presented in this paper.

2. Experimental method

The set up of the experiment and the experimental method of testing for the viscosity-density correlation are described in this section.

2.1. Rig description

The basic experimental rig setup (Figure 1, [10]) consists of two tanks, henceforth referred to as the jacketed tank and the hot water tank, respectively. The jacketed tank is fitted with an external jacket in which hot water is re-circulated. The jacketed tank also contains the milk concentrate in an inner tank. The hot water tank is a temperature controlled heat source, providing heated water which is circulated through the jacket of jacketed tank.

A Hydro-motion vibrational viscometer (XL/7) is inserted into the jacketed tank. This allows the viscosity of the milk concentrate to be measured at this point. Afterwards, the milk concentrate from the jacketed tank is passed though a micro-motion coriolis meter (DL65) allowing the density and flow rate of the system to be measured. The milk concentrate is then returned to the jacketed tank.

In order to control the milk concentrate temperature, a control valve is fitted on the water line from the hot water tank to the jacketed tank. The temperature of the jacketed tank is controlled by a control valve which adjusts the amount of heated water allowed to enter the jacket of the jacketed tank. Inside the hot water tank, a 3kW heating element is fitted along with a PID temperature controller and temperature sensor to regulate the water temperature.

Temperature is monitored at several places in the rig, including the hot water inlet to the Jacketed tank, the Jacketed tank hot water outlet, inside Hot water tank, at the micro-motion coriolis meter and inside the Jacketed tank. Other variables such as flow and density are monitored at the micro-motion coriolis meter within the jacketed tank.

2.2. Operating Procedure

The heating element is turned on to heat the water to a desired temperature. A controller is fitted to the heating element which cuts power to the element when the water reaches the desired temperature. For a typical run, the controller is set to cut out at 80 °C to maintain heated water to the rig.

Whole milk powder was used to prepare samples of milk concentrate with total solid contents of 35%, 40% and 45%. Once prepared, the sample is poured into the hot water jacketed tank for viscosity testing and recirculation. Once the jacketed tank is loaded, the pump was switched on to start to re-circulate of milk concentrate within the milk loop; the pump from hot water loop is also started to heat up the sample to the desired temperature.

From offline experimental results to minimise protein denaturation and structure change, the operating temperature for pilot plant is limited to 70°C. Structure change was observed above 80°C and it has been reported in literature that excessive whey denaturisation could not be avoided if it is operating above approximately 71°C [11].

Measurements were made during both heating and cooling of the concentrate to investigate whether the thermal history is an important factor in viscosity measurement. Data was collected during heating and cooling in the following temperature runs: 40 °C, 50 °C, 60°C and 70 °C. Other than at the maximum temperature of 70 °C, results were all collected twice. These two sets of results were then compared to check for consistency of viscosity measurements.

At the end of the experimental run, hot water is used to clean the rig by re-circulating at 70°C for 10 minutes. This is followed by a Clean In Place (CIP) procedure where caustic and acid clearing was performed.

3.   Results

3.1. Temperature sensitivity

Figure 2 presents the viscosity for both the heating and cooling cycles for whole milk concentrates of 35%,40% and 45% TS contents. The viscosity value was taken to be the average viscosity reading over a steady state of period of 10 minutes. Fluids were tested for both heating and cooling cycles to see if the fluid was still the same.

The results in **Figure 2** show that there is a strong downward trend, in which the viscosity of the fluid decreases as the temperature increases. The milk concentrate behaviour appears to observe the same trend when heating and cooling. This suggests that the milk concentrate is thermally stable over the experimental range.

3.2. Viscosity density relationship

The viscosity –density curves and correlations at temperatures between 40°C to 70°C are presented in Figure 3 and Table 1, in which we can see that there is a clear upward trend as expected.

4.   Discussion

The method presented in this paper offers a potential way of formulating a viscosity-density correlation. The fluid appears to be thermally stable within the experimental conditions. There appears to be no difference in viscosity after the fluid is heated and cooled down.

Viscosity and density correlations were able to be identified, an increase-increase relationship was found at all temperatures tested. This relationship is in agreement with expectation and literature. However, the slopes of the correlations are very shallow (0.0041-0.0090 $g.cm^{-3}/cP$) and there is some scatter in the data (R^2 is 0.77-0.93). This indicates that further study is needed, either with respect to improving these correlations or researching direct on-line viscosity measurement.

The method presented in this paper offers a potential way of testing for viscosity-density correlation which is useful for the formulation of a process model for advanced process control. In this experiment, reconstituted milk was used instead of fresh milk concentrate. This is expected to have an influence on the viscosity value of the milk concentrate, as the milk powder is exposed to high temperature at the spray drying chamber and this can cause protein denaturisation and hence viscosity change.

5.   Conclusions and Future work

The correlations developed by this method are empirical, and hence each different product specification (milk composition) needs to be tested if it is to be used for the model. It is expected that milk composition will affect the viscosity-density correlation. Further development is required in the search of density-viscosity correlation for milk. The approach presented above offers a way of identifying this correlation in steady state. However, for process control purposes, a dynamic correlation must be identified. This implies that the density and temperature needs to be changed and the response of viscosity needs to be observed.

Currently the sample used is reconstituted milk. It is expected that there is a difference in viscosity of milk concentrate between fresh milk concentrate and reconstituted milk concentrate. As this correlation is purely empirical correlation and highly possible site dependant. It is important perform the experiment with the feed from the particular plant.

In conclusion, this paper presents a proposed method for formulating a viscosity-density relation via an empirical approach. This correlation is potentially useful for the development of predictive control for milk concentrate when formulated, but requires further study and also testing on a specific site.

References

1.   Kelly, P.M., Innovation in milk powder technology. 2006. 59(2): p. 70-75.
2.   Setnes, M. and R. Babuska, Fuzzy decision support for the control of detergent production. International Journal of Adaptive Control and Signal Processing, 2001. 15(8): p. 769-785.
3.   Chen, X.D., Towards a comprehensive model based control of milk drying processes. Drying Technology, 1994. 12(5): p. 1105-1130.
4.   Govaerts, R., Johnson, A., Crezee, R., Reyman, G. and Swinkels, P.L.S., Control of an industrial spray drying unit. Control Engineering Practice, 1994. 2(1): p. 69-85.

5.   Allen, R.M. and H.H.C. Bakker, Spray dryer control based on online particle size analysis. Chemical Engineering Research and Design, 1994. 72(A2): p. 251-254.

6.   Zaror, C.A. and J.R. Perez-Correa, Model based control of centrifugal atomizer spray drying. Food Control, 1991. 2(3): p. 170-175.

7.   Clement, K.H., et al., On the dynamic behaviour of spray dryers. Chemical Engineering Research & Design, 1991. 69(3): p. 245-252.

8.   O'Callaghan, D. and P. Cunningham, Modern process control techniques in the production of dried milk products - A review. Lait, 2005. 85(4-5): p. 335-342.

9.   Schuck, P., Mejean, S., Dolivet, A., Beacher, E. and Famelart, M.-H., Pump amperage: A new method for monitoring viscosity of dairy concentrates before spray drying. Lait, 2005. 85(4-5): p. 361-367.

10. De Souza, G.F., Lin T.-I.; Young, B.R., Dairy Fluid Viscosity Measurement and Process Control, Proceedngs of the International Conference of Agricultural Engineering, Iguazu Falls, Brazil, 2008.11.

11. O'Sullivan, A.C., Whey protein denaturation in heat processing of milk and dairy products. Journal of the Society of Dairy Technology, 1971. 24(1): p. 45-53.
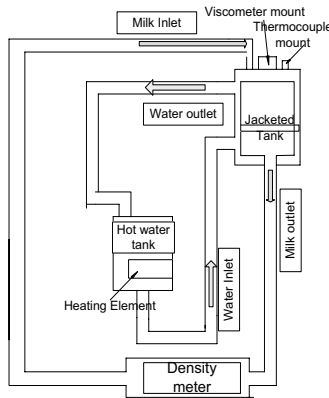
Figure 1. Schematic of the jacketed tank and hot water tank, adapted from [10]
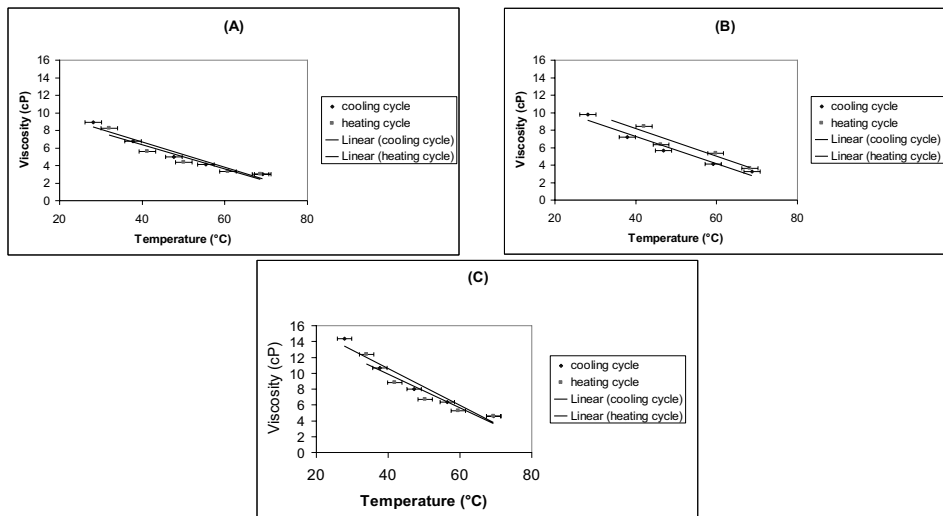
Figure 2. Viscosity versus Temperature for Skim Milk Concentrate for Heating/Cooling cycles, (A) At 35% TS, (B) At 40% TS, and (C) At TS 45% TS

Table 1. Density-viscosity correlation for whole milk between 35% to 45% total solids, where y= density (gcm^{-3}) and x= viscosity (cP)

Temperature (°C)	Correlation	R^2
40	y = 0.0041x + 1.0437	0.77
50	y = 0.0067x + 1.0324	0.93
60	y = 0.0064x + 1.0362	0.81
70	y = 0.0090x + 1.0259	0.87
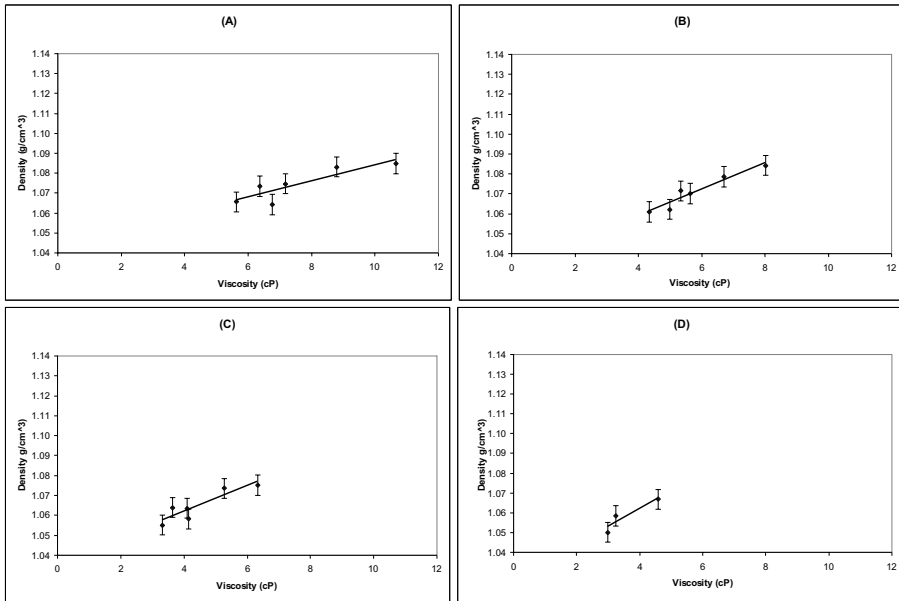
Figure 3. Viscosity-Density correlation for whole and skim milk over temperatures of 40 °C to 70°C: (A) At 40 °C, (B) At 50°C, (C) At 60 °C, and (D) At 70°C

1377

Learning to Repair Plans and Schedules Using a Relational (Deictic) Representation

Jorge Palombarini[a], Ernesto Martínez[b]

[a]UTN-Fac. Reg. V. María, Av. Universidad 450, Villa María 5900, Argentina
[b]INGAR (CONICET-UTN), Avellaneda 3657, Santa Fe S3002 GJC, Argentina

Abstract

Unplanned and abnormal events may have a significant impact in the feasibility of plans and schedules which requires to repair them 'on-the-fly' to guarantee due date compliance of orders-in-progress and negotiating delivery conditions for new orders. In this work, a repair-based rescheduling approach based on the integration of intensive simulations with logical and relational reinforcement learning is proposed. Based on a deictic representation of schedule states a number of repair operators have been designed to guide the search for a goal state. The knowledge generated via simulation is encoded in a relational regressión tree for the Q-value function defining the utility of applying a given repair operator at a given schedule state. A prototype implementation is discussed using a representative example of 3 batch extruders processing orders for 4 different products. The learning curve for the problem of inserting a new order vividly illustrates the advantages of logical and relational learning in rescheduling.

Keywords: rescheduling, artificial intelligence, batch plants, relational modeling.

1. Introduction

Fast rescheduling in real-time is mandatory to account for unplanned and abnormal events including arrival of new orders, equipment breakdown, reprocessing and raw material delays (Vieira et al, 2003). Reactivity and responsiveness is a key issue in any rescheduling strategy which makes critical the capability of generating and representing knowledge about heuristics for repair-based scheduling using case-based reasoning (Miyashita, 2000). One such example is the **CABINS** framework for case-based rescheduling proposed by Miyashita and Sycara (1994) that heaviliy resorts to human experts. Along similar ideas another important work in the field of the so-called *intelligent scheduling* techniques are contributions by Zweben et al. (1994). In this work the novelty lies in integrating intensive simulations with automatic learning to overcome the issue of human experts for domain-specific scheduling problems. A novel methodology which combines a deictic representation of schedules states with relational reinforcement learning is proposed to develop a near-optimal policy for interactive scheduling repair bearing in mind different goals and scenarios. To this aim, learning to repair schedules based on simulation is undertaken using two general-purpose algorithms already available: TILDE and RRL (Džeroski et al, 2001; De Raedt, 2008).

2. Repair-based (re)scheduling

Fig. 1 depicts the repair-based optimization architecture where search control knowledge about repair operator selection is acquired through reinforcements using a schedule state simulator. In the simulation environment an instance of the schedule is

interactively modified by the learning agent using a set of repair operators until a goal is achieved or the impossibility of repairing the schedule is accepted. In each interaction, the learning agent receives information from the schedule situation or state s and selects a repair operator to be applied to the current schedule as an action a. The resulting quality of a schedule after the repair operator has been applied is evaluated using the simulation environment via an objective or reward function $r(s)$. The learning agent then updates its action-value function $Q(s,a)$ that estimates the value or utility of resorting to the chosen repair operator a in a given schedule state s. Such an update is made using a reinforcement learning algorithm (Sutton and Barto, 1998) such as the well-known Q-learning rule. By accumulating enough experiences over many simulated interactions the agent is able to learn an optimal policy for choosing the best repair operator at each schedule state. The main issue for learning is then how schedules states and actions must be represented for knowledge acquisition and iterative revision.
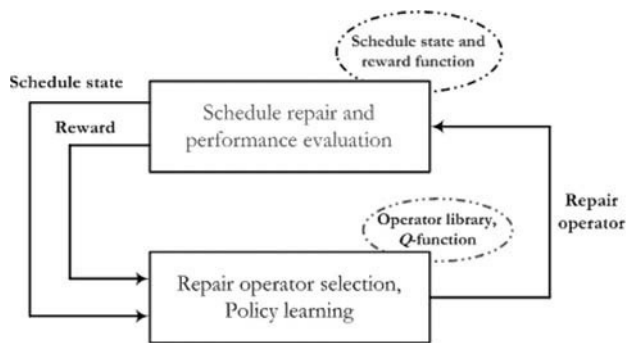
Fig. 1. Knowledge acquisition for schedule repair using reinforcement learning

For repairing a plan or schedule, the agent is given a *goal-state $S :\rightarrow \{true, false\}$* defining which states in the planning world are target states. The objective of a repair task is phrased as: Given a starting state schedule s_1, find a sequence of repair operators $a_1, a_2,..., a_n$ with $a_i \in A$ such that the $S \equiv true$. Usually a precondition function is used to specify which subset of repair operators can be applied at each state of the planning world to account for resource and precedence constraints. Also, a reward function is used by a learning algorithm to develop a repair policy using simulations (Martínez, 1999). The key issue is though how a schedule state must be best represented so that the repair policy can be applied to unseen schedule states more effectively. Propositional representations are not adequate for learning in open planning worlds defined by tasks, their properties and their relations to other tasks and resources. A *deictic* representation dealing with the varying number of tasks in the planning world by defining a focal point for referencing tasks is proposed here as a much powerful alternative. In a deictic representation, the rest of the planning world is then defined in relation to that focal point as it is shown in Fig. 2. These local repair operators move the position of a task alone, however due to the ripple effects caused by tight resource-sharing constraints other tasks may need to be moved which is not desirable. Whenever the goal-state for the schedule cannot be achieved using primitive repair operators more elaborated macro-operators can be used to implement a combination of basic repair operators such as `task-swapping`, `batch-split` or `batch-merge` until the goal state (e.g. order insertion without delaying other orders) is achieved.

3. Relational reinforcement learning (RRL)

RRL algorithms are concerned with reinforcement learning in domains that exhibit structural properties and in which different kinds of related objects such as tasks and resources exist (Džeroski et al, 2001; De Raedt, 2008). These kind of domains are usually characterized by a very large and possibly unbounded number of different possible states and actions. In this kind of environments, most traditional reinforcement learning techniques break down. One reason why traditional RL fails is that it stores the learned Q-values explicitly in an state-action table, with one value for each possible combination of states and actions. Rather than using an explicit state−action Q-table, RRL stores the Q-values in a logical regression tree (Blockeel and De Raedt, 1998).
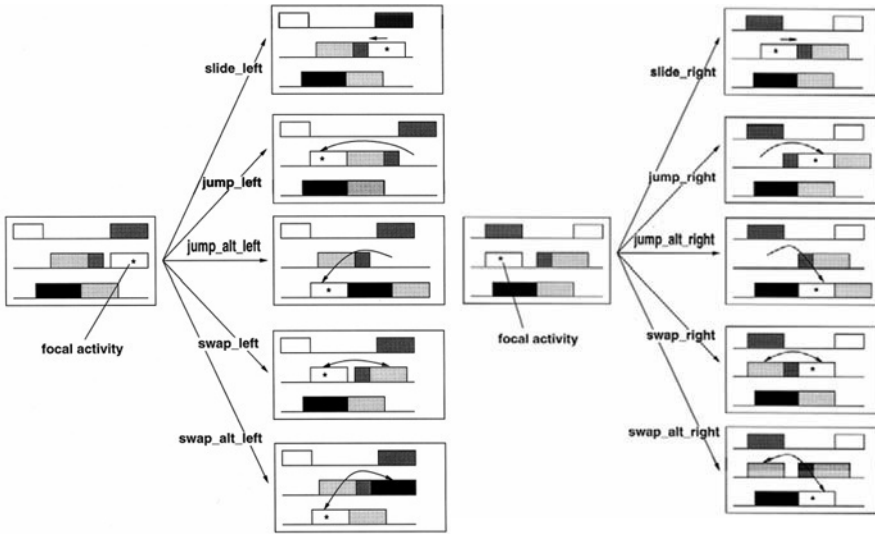
Fig. 2. Deictic representations of repair operators

In RRL states are represented as sets of first-order logical facts, and the learning algorithm can only see one state at a time. Actions are also represented relationally as predicates describing the action as a relationship between one or more variables. Because of the relational representation of states and actions and the inductive logic programming component of the RRL algorithm, there must exist some body of background knowledge which is generally true for the entire domain. After the Q-function hypothesis has been initialized, the RRL algorithm starts running learning episodes like any standard Q-learning algorithm (Sutton and Barto, 1998; Džeroski et al, 2001). During each learning episode, all the encountered states and the selected actions are stored, together with the rewards related to each encountered (state, action) pair. At the end of each episode, when the system encounters a *goal* state, it uses reward back-propagation and the current Q-function approximation to compute the appropriate Q-value approximation for each encountered (state, action) pair in the episode. The algorithm then presents the set of (*state*, *action*, *qvalue*) triplets to a relational regression engine, which will use this set of *Examples* set to update the current regression tree of the Q-function, and then the algorithm continues executing the next learning episode.

Initialize the Q-function hypothesis $\hat{Q}_0$
$e \leftarrow 0$
repeat
   Examples $\leftarrow \varnothing$
   Generate a starting schedule state s_0
   $i \leftarrow 0$
   repeat
      choose a repair operator a_i at s_i using a policy (e.g., ε-greedy) based on the current
      hypothesis $\hat{Q}_e$ implement operator a_i, observe r_i and the resulting schedule s_{i+1}
      $i \leftarrow i+1$
   until schedule state s_i is a *goal* state
   for $j = i-1$ to 0 **do**
      generate example $x = (s_j, a_j, \hat{q}_j)$, where $\hat{q}_j \leftarrow r_j + \gamma \max_a \hat{Q}_e(s_{j+1}, a)$
      Examples $\leftarrow$ *Examples* $\cup \{x\}$
   end for
   Update $\hat{Q}_e$ to $\hat{Q}_{e+1}$ using *Examples* and a relational regression algorithm (e.g. TG in Fig. 4)
until no more learning episodes

Fig, 3. A RRL algorithm for learning to repair schedules through intensive simulations

The TG algorithm described in Fig. 4 is a relational regression algorithm that have been developed for policy representation in logical and relational learning (De Raedt, 2008). This incremental first order regression tree algorithm is used here for accumulating simulated experience in a compact representation of a repair-based policy based on Q-values for the repair operators available at the state s. Fig. 5 gives a small example of a first order regression tree for the Q-value function in a task (re) scheduling world to react to events and disturbances.

//initialize by creating a tree with a single leaf with empty
statistics
for each learning example that becomes available **do**
   sort the example down the tree using the tests of the
   internal nodes **until** it reaches a leaf
   update the *Q-value* in the leaf according to the new
   example
  if the statistics in the leaf indicate that a new split is
  needed **then**
   generate an internal node using the indicated test
   grow 2 new leafs with empty Q statistics
  end if
end for

Fig. 4. TG algorithm for relational induction   Fig. 5. A relational regression tree
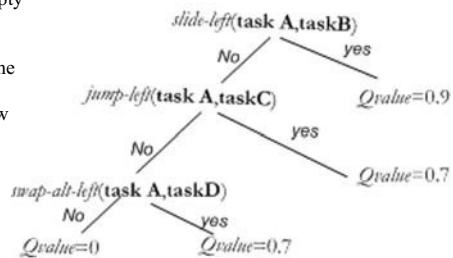
4. Example

A small example problem proposed by Musier and Evans (1989) is considered to illustrate the use of repair operators for batch plant rescheduling. The plant is made up of 3 semicontinuous extruders which process customer orders for 4 products. Processing rates and cleanout requirements are detailed in the quoted reference.
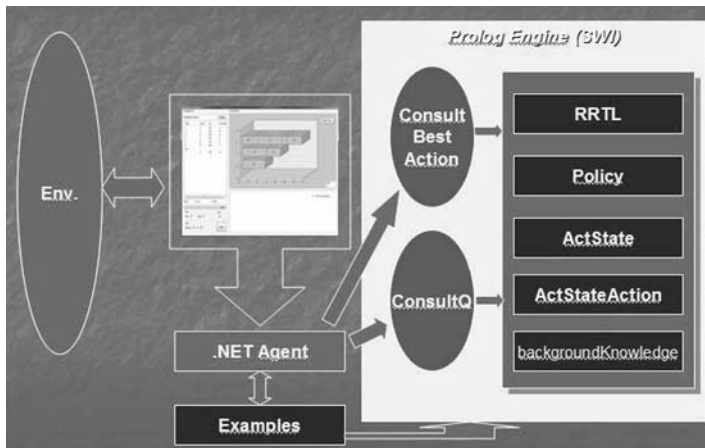
Fig. 6. SWI Prolog engine for implementing RRL in repair-based (re)scheduling

The prototype application has been implemented in Visual Basic.NET 2005 Development Framework 2.0 SP2 and SWI Prolog 5.6.61 running under Windows Vista. Also, the **TILDE** and **RRL** modules from The **ACE Datamining System** developed by the Machine Learning group at the University of Leuven have been used. Let's discuss only the specific situation where there exist 10 already scheduled orders in the plant and a new order #11 must be inserted so that the Total Tardiness (TT) in the schedule is minimized. Data for scheduled orders and the new order are given in Table 1. En each episode, a random schedule state for orders #1 through #10 is generated and a random insertion attempted for the new order which in turn serves as the focal point for defining repair operators. Also, a goal state for the resulting schedule is stated in terms of the TT, e.g. of 5% increase. Background knowledge such as "the number of orders scheduled for extruder #3 is larger than the number for extruder #2" is provided to speed up learning. In Fig. 7, the learning curve for the new order insertion rescheduling event is shown. Training episodes comprise of a number of steps, each one corresponding to the application of one local repair operator. After 15 training episodes only 3 repair steps are required on average to insert the 11th order. Fig. 8 provides an example of the optimal sequence of repair operators from the schedule in Fig. 8(a).

5. Final comments

A novel approach for simulation-based development of a relational policy for automatic repair of plans and schedules using reinforcement learning has been proposed. The policy allows generating a sequence of deictic (local) repair operators to achieve a goal such as order insertion with minimum tardiness based on deictic representations.

6. Acknowledgements

References

H. Blockeel, L. De Raedt, 1998, Top-down induction of first order logical decision trees. *Artificial Intelligence*, 101, 1-2, 285–297.

K. Driessens, 2004, *Relational reinforcement learning*, Ph.D. dissertation, Department of
    Computer Science, K.U.Leuven, Leuven, Belgium..
S. Džeroski, L. De Raedt, K. Driessens, 2001, Relational reinforcement learning, *Machine
    Learning*, 43, 1/2, 7−52.
L. De Raedt, 2008, *Logical and relational learning*, Springer-Verlag, Berlin, Germany.
E. C. Martinez, 1999, Solving batch process scheduling/planning tasks using reinforcement
    learning, *Computers and Chemical Engineering*, 23, S527-531.
K. Miyashita, 2000, Learning scheduling control through reinforcements, *International.
    Transactionsin Operational Research* (Pergamon Press), 7, 125-138.
K Miyashita, K. Sycara, 1994, **CABINS**: a framework of knowledge acquisition and iterative
    revision for schedule improvement and iterative repair, *Artificial Intelligence*, 76, 377-426.
R. Musier, L. Evans, 1989, An approximate method for the production scheduling of industrial
    batch processes with parallel units, *Computers chem Engng*, 13, 1-2, 229-238.
R. Sutton, A. Barto, 1998, *Reinforcement Learning: An Introduction*, MIT Press, Boston, MA.
G. Vieira, J. Herrmann, E. Lin, 2003, Rescheduling manufacturing systems: a framework of
    strategies, policies and methods, *J. of Scheduling*, 6, 39-62,
M. Zweben, E. Davis, B. Doun, M. Deale, 1993, Iterative repair of scheduling and rescheduling,
    IEEE. Trans. Syst. Man Cybern., 23, 1588-1596.
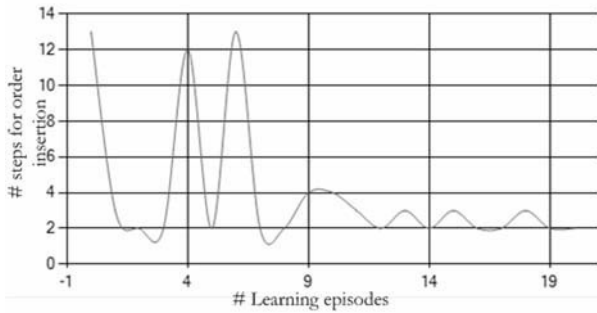
Fig. 7. Learning curve for the repair-based order insertion example

Table I. Data for the new order insertion example

#	Product	Size (Kg)	DD (h)	#	Product	Size (Kg)	DD (h)
1	A	300	6	6	B	600	5
2	B	300	5	7	A	400	6
3	C	700	3	8	B	500	12
4	D	100	2	9	C	700	17
5	D	700	10	10	C	300	8
11	A	150	10				

| (a) | (b) | (c) |
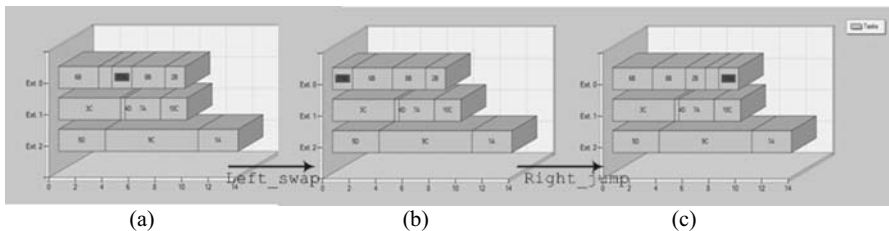
Fig. 8. Before inserting order #11 (in red) the is TT=17.5. The goal is to repair the schedule in (a)
having a TT=20 h so that TT increase is minimized; (b) After a `Left_swap` (b) TT=19 h.; After
a `Rigth_jump` in (c) TT=17.50, meaning an insertion is found without any increase in TT.

1383

Development of Multi-Modal Control Programs for Continuous-Discrete Process Supervision

Mariano De Paula, Ernesto Martínez

aINGAR (CONICET-UTN), Avellaneda 3657, Santa Fe S3002 GJC, Argentina

Abstract

Continuos-discrete processes are characterized by the strong coupling of continuos state dynamics of recipe-driven discontinuos operations and the discrete state dynamics of logic-based controllers acting at the interface with continuos processes. Multi-modal control is a promising design tool for supervisory control of these type of systems by resorting to sequences of control modes each one comprising of a purposeful feedback law and a stopping condition. In this paper the problem of developing multi-modal control programs from a given mode alphabet using data from alternative disturbance scenarios is addressed. A case study related to maximizing the average productivity of a hybrid chemical plant through multi-modal control of a buffer tank is presented.

Keywords: Batch control, hybrid plant, supervisory control, mode-switching control.

1. Introduction

There has been a lot of interest in modelling and control of hybrid dynamical systems in recent years, motivated by a growing number of hybrid chemical plants involving tight interaction between continuous variables and discrete events/decisions (Engell, *et al.*, 2000). A multi-modal system is a system whose dynamics switches among finitely many possibilities to achieve a desired goal or behavior. These switches can be in response to an occurrence of a specific event or a controlled decision (Liberzon, 2003; Mehta and Egerstedt, 2006). Optimal control of multi-modal dynamic systems is of paramount importance for designing a supervisory control structure (Koutsoukos, *et al.*, 2000). In this context, a mode σ is specified as a triple $(\kappa(\bullet), \xi(\bullet), T)$, where $\kappa(\bullet)$ is a control law (closed-loop control policy) seeking to achieve a given sub-goal or to deploy a certain behavior, $\xi(\bullet)$ is the interrupt function mapping process observations to the set $\{0,1\}$ and T is the time interval over which the mode σ is active. The key issue in the design of a supervisory control system is to find the optimal concatenation of modes in a finite-length mode string π based on the rewards observed after each mode transition occurs.

2. Multi-modal control programs

Suppose the state x of a dynamic system follows:

$$\frac{dx}{dt} = f(x, u(t), z(t)), \ x \in X = R^n, u \in U = R^m \tag{1}$$

where $z(t)$ is a time-varying measurable disturbance which evolves arbitrarily as

$$\frac{dz}{dt} = g(z,t), \ z \in Z = R^d \tag{2}$$

If at a given time τ_0, where the state is $x(\tau_0)$ and the disturbance is $z(\tau_0)$, the system receives the input string $\pi = \{(\kappa_1, \xi_1), \ldots, (\kappa_q, \xi_q)\}$ which gives rise to the state transition

$$\delta(\sigma, x(\tau_0), z(\tau_0)) = x(\tau_0) + \int_{\tau_0}^{\tau_1} f(x(t), u(t), z(t)dt + \ldots + \int_{\tau_{q-1}}^{\tau_q} f(x(t), u(t), z(t)dt \tag{3}$$

If the length of the mode sequence π is bounded, only a finite set of states are reachable. Hence, by resorting to control programs with maximum length N a quantization of the system state results. Based on this multi-modal quantization a Lebesgue-sampled finite state machine $(X_N^Q, \Sigma, \tilde{\delta}, \tilde{x}_0, \tilde{z}_0)$, where Q stands for the value function (Sutton and Barto, 1998) and the state transition $\tilde{\delta}$ is defined as follows

$$\tilde{x}_{k+1} = \tilde{\delta}(\tilde{x}_k, \tilde{z}_k, \sigma_k) = \delta(\sigma, x(\tau_0), z(\tau_0)), k = 0,1,2,\ldots., \ x_0 = x(\tau_0), z_0 = z(\tau_0), \tag{4}$$

The discrete state space X_N^Q is given by the set of all states that are reachable from $\chi_0 = (x_0, z_0)$ using a sequence of modes $\sigma \in \Sigma$ of length less than or equal to N. The issue of concern here is finding a sequence of control-interrupt pairs that maximizes a cumulative reward for such state transitions. Based on the multi-modal quantization X_N^Q of the state space and the transition dynamics of the Lebesgue-sampled FSA, reinforcement learning (Sutton and Barto, 1998) can be used to learn the Q-values for a given reward function. It is worth noting that both state exploration and disturbance scenarios are required to discover which is the best mode $\sigma*$ for each pair $\chi \in X_N^Q$.

$\aleph := \{\tilde{\chi}_0, \delta(\tilde{\chi}_0, \sigma)\}$; *step* $(\tilde{\chi}_0) := 0$;

step $\tilde{\delta}(\tilde{\chi}_0, \sigma) := 1, \ \forall \sigma \in \Sigma$

$k := 1$; *index for counting visits to state-disturbance pairs* $\chi \in \aleph$

$Q_k(\tilde{\chi}, \sigma) := const, \forall \chi \in \aleph, \sigma \in \Sigma$

repeat

    $k = k+1$

    $(\tilde{\chi}) := rand(\chi \in \aleph | step(\chi) < N)$

    $\sigma := rand(\Sigma)$

    $\tilde{\chi}' := \tilde{\delta}(\tilde{\chi}, \sigma)$

    if $\tilde{\chi}' \notin \aleph$ **then**

        $step(\tilde{\chi}') = step(\tilde{\chi}) + 1$

        $\aleph := \aleph \cup \tilde{\chi}'$

        $Q_k(\tilde{\chi}', \sigma) := const, \forall \chi \in \aleph, \sigma \in \Sigma$

    end if

    $Q_{k+1}(\tilde{\chi}, \sigma) := Q_k(\tilde{\chi}, \sigma) +$

    $\alpha_k \{r(\tilde{\chi}, \sigma) + \gamma \max_{\sigma' \in \Sigma} [Q_{k-1}(\tilde{\chi}', \sigma') - Q_{k-1}(\tilde{\chi}, \sigma)]\}$

until $\mod(k, L) = 0$ and $\left| Q_k(\tilde{\chi}, \sigma) - Q_{k-L}(\tilde{\chi}, \sigma) \right| < \varepsilon, \forall \chi \in \aleph, \sigma \in \Sigma$

$X_N^Q = \aleph$

Fig. 1. *Q*-learning algorithm for multi-modal control programs

In the Q-learning algorithm of Fig. 1, the quantized state-disturbance space X_N^Q is assumed to be unknown initially and we thus begin learning/exploring from selected states/disturbances pairs $\tilde{\chi}_0 = (\tilde{x}_0, \tilde{z}_0)$ and simulating transitions toward all the states which are reachable using the first mode of all control programs $\sigma \in \Sigma$. At each iteration of the learning process, a state-disturbance pair is randomly chosen from the set of known states-disturbance pairs and a control mode from the set Σ is applied to observe a state transition to $\tilde{x}^+$; $r(\tilde{\chi}', \sigma)$ is the reward obtained for the transition observed once the execution of mode σ is finished.

In the algorithm, the function $step(\tilde{x})$ represents the length of the shortest control program used so far to reach the state $\tilde{x}'$ from the initial state $\tilde{x}_0$. As a result, only states that are reachable from $\tilde{x}_0$ using mode strings of length less than or equal to N are explored, i.e. $\tilde{x}' \subset \tilde{\chi}' \in X_N^Q$ is guaranteed. As the transition to the next state $\tilde{x}'$ is observed it is necessary to determine if it is in the neighborhood of a previously visited state-disturbance pair $\tilde{\chi}$. If not the pair $\tilde{\chi}$ is added to the known state space-disturbance, the coresponding number of steps is increased by 1 and a new entry to Q-table is added. It is worth noting than when $\tilde{\chi}'$ is in the the neighborhood of a previously visited state the Q-values of all states-disturbances pairs in the neighborhood are updated. Exploration of the state-disturbance space and updates of the Q-table (value function) continues in this manner until the Q-table is stationary. Stopping conditions are imposed so as to guarantee that sufficiently many state-disturbance pairs are visited and all control modes $\sigma \in \Sigma$ are tried enough times for the Q-table to converge while the learning rate α is increasingly reduced.

3. Case study

3.1. Hybrid chemical plants

Buffer tanks play a key role in hybrid chemical plants such as the *Solvay* PVC production line (Melas, 2003) or sugar milling plants (Ghaeli et al., 2008) to smooth the interface between the batch and continuos sectors in the plant. To maximize the average plant productivity a buffer tank is used as shown in Fig. 2. This type of processes are generally made up of several parallel working units (e.g., batch reactors or crystallizers) "sharing" common resources, typically utilities (e.g. cold water, power or steam). Batch reactors or crystallizers often operates by cyclically passing through a sequence of phases, namely loading of raw material, cooling, heating, reaction, etc.. Resource sharing by different process unit may increase the duration of these tasks in the recipe which will alter the whole schedule of batches in the plant. Also, as batch units discharge their final product in a buffer tank used to transfer it continuously to the continuous units in the plant (centrifuges, dryers, etc.) such delays may drastically the pattern of the inflow to the tank. There also exist some "waiting" times for a batch which can be used to compensate for deviations in the original schedule but at the cost of a productivity loss.

Maximization of average outflow from the tank is the main objective for productivity optimization in this hybrid chemical plant. However, another important component of the desired tank operation is that its outflow rate F_{out} must be changed smoothly despite significant variations in its inflow rate. To do so the tank capacity must be managed properly by allowing the tank level to vary within its minimum and maximum limits. Also, the outflow rate F_{out} is constrained by minimum and maximum values that must

be enforced due to throughput limitations downstream in the continuos processing sector. It is worth noting that in hybrid chemical plants the main role of the buffer tank is handling intermitent inflow rates which prevents resorting to averaging level control techniques (McDonald, et al., 1986). Also, for this type of processes the inflow rate variability pattern may be the result of significant deviations from periodic or optimal schedules due to resouce sharing, e.g. utilities (Simeonova, 2008).

How does multi-modal control of the buffer tank fits in the problem of productivity optimization of the hybrid plant? Let's take a quick look to what may be expected from the control system to elaborate on some intuitive insights in the search for a well-posed solution. Firstly, it is important that feedback laws allows that at any time the outflow rate must be a monotonic function of the tank level. This selection introduces a self-regulation property into the multi-modal control strategy. Also, modes must be designed so that despite peaks and valleys in inflow rate mode switching makes possible to maintain the average outflow rate as high as possible throughout. Thus, different modes should provide different degree of aggressiveness through their feedback laws when seeking a sub-goal (accumulate or drain-off inventory). Finally, rewards obtained after each mode application must be defined bearing in mind productivity maximization, e.g. the mean value of the outflow rate over the time inteval T over which the mode was applied. Large negative rewards are given to modes whose applications results in less than a minimum outflow rate or tank overflowing. To avoid undesirable disruptions in downstream units once a mode change occurrs the corresponding feedback laws must be applied for a minimum number of time intervals, say 5 min. This avoids excessive mode switching without compromising mode sequence optimality.

3.2. Tank inflow data and control modes

As a representative example of the application of the algorithm in Fig. 1 to a buffer tank, let's assume the buffer tank has a maximum height of 1 m which is equal to its diameter. The inflow rate variability is modeled here as the time-series generated using the well-known Mackey-Glass differential delay equation using random initial values

$$\frac{dF_{in}}{dt} = \frac{0.2 F_{in}(t-30)}{1 + \left(F_{in}(t-30)\right)^{10}} - 0.9 F_{in}(t) \ [\text{m}^3/\text{h}] \tag{5}$$

Time series generated using Eq. (5) are very sensitive to initial conditions while exhibiting a non-periodic/non-convergent pattern as can be seen from portions of the series shown in Fig. 3. This type of inflow variability may look far from real but it will serve as an extreme benchmark for developing multi-modal control programs using the algorithm of Fig. 1 to handle disturbance scenarios where significant variations in the upstream parallel unit discharge schedule are present. Modes are designed so as to achieve either the sub-goal *accumulate* (A1 and A2) or the sub-goal *drain-off* (D1 and D2) using for both the following the two simple feedback laws:

$$\kappa_1 : F_{out}(t) = \frac{0.4.[h(t)]^{0.5}}{1.2 - h(t)} - 0.17 \quad \text{and} \quad \kappa_2 : F_{out}(t) = 0.934.h(t) - 0.15 \tag{6}$$

where h is the tank level (see Fig. 4). The built-in smoothness of the feedback laws in (6) is important to reduce the amplitude of outflow rate variations as much as possible. Once the optimal sequence of modes is found feedback laws and their parameters can be optimized using a variational formulation with the objective of lowering the complexity

of π and smoothing the outflow rate. Stopping conditions are defined based on the failure to achieve the corresponding sub-goal over time. For example, for any accumulation (drain-off) mode its stopping event triggers the corresponding variable change from $\xi=0$ to $\xi=1$ as soon as the level stop increasing (decreasing). For describing the disturbance scenarios for inflow rate a 2-entries vector $z(t)=(\rho(t),F_{in}(t))$ is used where $-1<\rho(t)<1$ is the local correlation coefficient calculated based on the current and 3 precedent values of $F_{in}(t)$. After each mode implementation, rewards given are calculated as the avergage outflow rate over the time interval between transitions as long as the tank does not overflow nor becomes empty. When an overflow or shut-down event occurrs a large negative reward is given to the state transition.
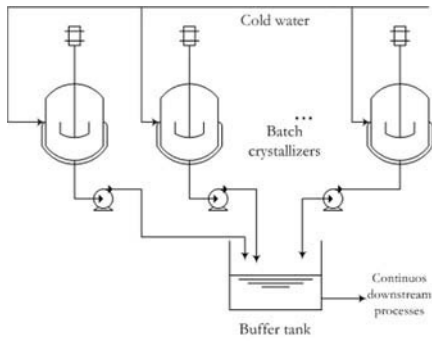
Fig. 2. Hybrid chemical plant   Fig. 3. Inflow rate variability pattern
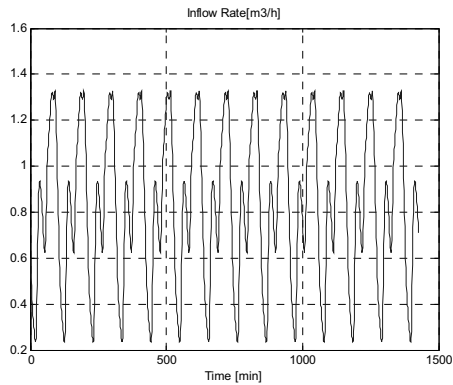
In Eq. (6) above the feedback laws used for both to accumulate and drain-off are given. As a result four different modes are available to maximize throuhgput whilst avoiding tank overflow or drownstream feeding interruptions. Results obtaned for some input disturbance scenarios are summarized in Fig. 4 through 7. It's noteworthy in Fig. 4 and 5 that despite abrupt changes in this scenario for the inflow rate (see Fig. 3) the multi-modal control based on the learned Q-values is able to maintain the outflow rate significantly high. Note that the tank is operated quite close the maximum level to maximize the average productivity whereas mode switching is used to avoid overflow. The corresponding control program is the mode sequence $\pi:$ (A1-D1-A1-D2).
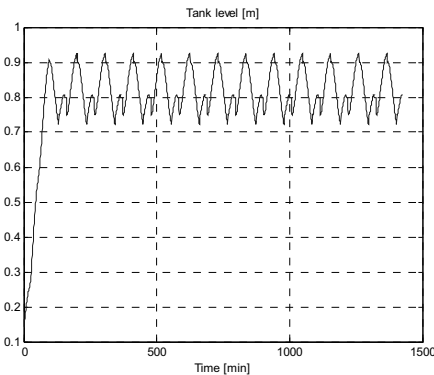
Fig. 4. Tank level under multi-modal control   Fig. 5. Outflow rate under multi-modal control
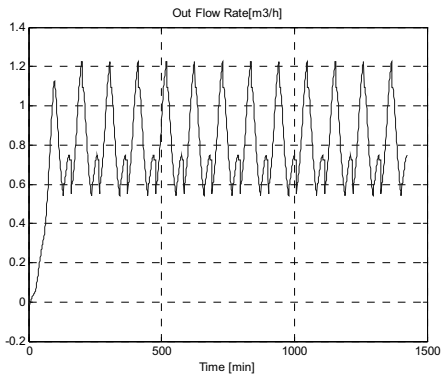
Finally, let's consider a rather different scenario for the inflow rate and tank situation when its initial level is very low. As shown in Fig. 6 and 7 the multi-modal control program restores the maximum productivity condition in the tank using only mode A1.
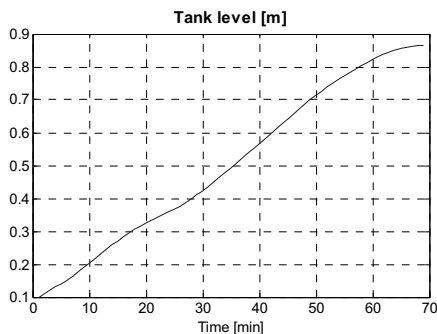
Fig. 6. Tank level recovery   Fig. 7. Outflow rate maximization
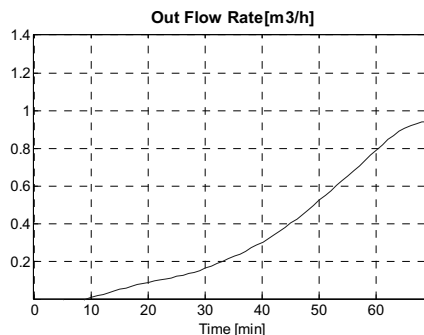
4. Final remarks

A new approach for automatic generation of a multi-modal control program based on simulations and reinforcement learning has been presented using buffer tank control for productivity maximiazation in a hybrid chemical plant. The main advantage of the proposed methodology is that if the inflow rate pattern changes the Q-function for state-modes in the program will change and the multi-modal control policy will be updated.

5. Acknowledgements

References

P. Barton, C. Lee, M. Yunt, 2006. Optimization of hybrid systems, Computers chem Engng, 30, 1576-1589.

S. Engell, S. Kowalewski, C. Schulz and O. Stursberg, 2000, Continuous-Discrete Interactions in Chemical Processing Plants, Proceedings of the IEEE 88, 1050-1068.

M. Ghaeli, P. Bahri, P. Lee, 2008, Scheduling of a mixed batch/continuous sugar milling plant using Petri nets, Computers chem Engng, 32, 580-589.

X. Koutsoukos, P. Antsaklis, J. Stiver and M. Lemmon, 2000, Supervisory Control of Hybrid Systems, Proceedings of the IEEE 88, 1026-1049.

D. Liberzon, 2003, Switching in Systems and Control, Systems & Control: Foundations & Applications, Birkhäuser Boston Inc., Boston, MA.

K. McDonald, T. McAvoy, and A. Tits, 1986, Optimal averaging level control, AIChE J., 32, 75–86.

T. Mehta and M. Egerstedt, 2006, An optimal control approach to mode generation in hybrid systems, Nonlinear Analysis 65, 963–983.

S. Melas, 2003, Pvc line predictive inventory control: a production rate optimization for a hybrid system, Solvay-Group report, 1–34.

I. Simeonova, 2008, On-line periodic scheduling of hybrid chemical plants with parallel production lines and shared resources, PhD Thesis, Université Catholique de Louvain.

R. Sutton and A. Barto, 1998, Reinforcement learning:An Introduction, The MIT Press, Cambridge, MA.

Advanced Temperature Tracking Control for High Quality Wines Using a Phenomenological Model

Oscar A. Ortiz[a], Martha D. Vallejo[b], Gustavo J. E. Scaglia[a], Carmen A. Mengual[a] and Pablo M. Aballay[a]

[a]*Instituto de Ingeniería Química, Universidad Nacional de San Juan, Av. Lib. San Martín Oeste 1109, J5400ARL San Juan, Argentina*
[b]*Instituto de Biotecnología, Universidad Nacional de San Juan, Av. Lib. San Martín Oeste 1109, J5400ARL San Juan, Argentina.*

Abstract

The fermentation in winemaking has high complexity due to interactions between cell biokinetics and bioreactor hydrodynamics. Hence, the new technological options to obtain high quality wines with outstanding organoleptic characteristics require more strict process monitoring and control based on rigorous models. This work presents an advanced temperature control system based on an improved non-isothermal phenomenological model that allows tracking complex temperature profiles to achieve optimal quality of wine. The controller has been performed in discrete-time, so that the current disturbance effect at the output is computed as the difference between the current measured value of the output and the predicted one. The obtained results are satisfactory for experimental data from literature.

Keywords: Dynamic inverse control**,** temperature tracking, wine fermentation.

1. Introduction

The growing customers demand for high quality wines and its marked preferences for outstanding wine organoleptic properties, presents new challenges for the winemaking. Hence, new technologies are recently emerging. Some of them include variable temperature trajectories throughout the fermentation. Indeed, there are strong evidences about the contribution of temperature to the release of some varietal aroma precursors and to the synthesis of volatile compounds by the wine yeast. Temperature affects metabolism and gene expression of the yeast, and is also an important variable in the maceration phase (extraction and diffusion of sugars, acids, aroma/taste precursors, etc., from the grape), that may be carried out in a pre-fermentation stage or during the fermentation. Temperature is seen as a fine tool to a better regulation of the fermentation progress, in accordance with a winemaking strategy depending on the desired wine (Torija *et al*., 2003; Molina *et al*., 2007; Romero Cascales, 2008). On the other hand, fermentation presents great interactions between cell biokinetics and bioreactor hydrodynamic conditions, which lead to models with non-linear and unsteady characteristics. Therefore, advanced optimization and control tools for monitoring and controlling strictly the process, are required. The implementation of advanced control strategies needs appropriate dynamic models and reliable on-line measurements (Henson, 2003). So, more rigorous and accurate models are necessary. They must be at the same time precise and simple, from a mathematical point of view, to be used in on-line control algorithms. In previous contributions, the authors have addressed isothermal

and non-isothermal first principles and hybrid models, and an improved isothermal phenomenological model with satisfactory capability to approximate the wine fermentation profiles (Vallejo *et al.*, 2005; Ortiz *et al.*, 2006; Aballay *et al.*, 2008; Scaglia *et al.*, 2009).

In this work, an advanced temperature control system for an enological batch alcoholic fermentation is presented. The control system use a non-isothermal phenomenological model that couples mass and energy balances to predict the main process state variables: viable cells, sugar and ethanol concentrations, carbon dioxide released, and the bioreactor temperature. The energy balance takes into account, the heat exchanged between the bioreactor and its cooling jacket and, the cellular growth rate dependence on temperature. Variations of specific heat and density in the bulk have been included in the model. Section 2 presents a brief description of the fermentation in winemaking and its modelling. Also, the non-isothermal model with considerations on each one of its components and parameters, as well as on its main properties is described in this section. In section 3, the dynamic model performance is evaluated by means of a wide performance analysis. The most frequently found disturbances in actual plants are applied on operating process variables and main model parameters and, the corresponding responses of the state process variables are shown. The control problem is addressed in section 4 where the main control objective has been posed on temperature, which is required to change according to the fermentation behavior and technological options, such as maceration or to promote enzymatic activities. The proposed control on temperature is based on the manipulation of cooling water flow rate through a dynamic inverse controller. The corresponding simulations of the controller response, with and without model parameters errors, are shown. Also, the errors between the control system response and the reference trajectory are analyzed. Section 5 presents the discussion of the main obtained results, analyzes the contribution of the work and outlines some topics to be addressed in future works.

2. Non - isothermal model

The grape juice fermentation is the winemaking heart. Since it is operated on batch mode, yeasts must get rapidly adapted to highly variable environmental conditions, which greatly difficult the kinetic predictions of the bioprocess. Numerous and complex physical, chemical and biological changes occur at this stage and fix on the final wine quality (Boulton *et al.*, 1996; Cramer *et al.*, 2002; Rosenfeld *et al.*, 2003). On the other hand, some outstanding organoleptic properties only are achieved tracking rigorously predefined temperature trajectories. For example, maceration at around 28-30ºC is generally required at the beginning of the fermentation, because enzymes are affected by the increasing ethanol concentration and better enzymatic activities are achieved at these "high" temperatures; 28ºC is appropriate for a fast start-up of the yeast multiplication, favoring the implantation of the desired population and inducing the synthesis of flowery related aroma compounds; temperatures between 23 and 18ºC are standard in winemaking and avoid ethanol and volatile compound losses; or definite stages at low temperature (near 15ºC) often stated to produce compounds related to fresh and fruity aromas in the wines, maintaining varietal aromas (Esti and Tamborra, 2006; Beltrán *et al.*, 2008). Significant but controllable and not stressing changes occur in the fermentative activity when temperature is slightly changed. Such control and monitoring requirements, establish each more time the availability of more accurate non-isothermal models. The reductive metabolic pathway S ➔ X + P + CO_2, characterizes the yeast population growth in anaerobic conditions. X, S, P and CO_2

correspond to viable cells, carbon source (glucose and fructose) and ethanol concentrations, and carbon dioxide released [kg m^{-3}], respectively. The metabolite accumulation in the extra cellular medium has been modelled by a set of ODE based on mass balances on X, S, P and CO_2 like in the isothermal model of Scaglia *et al.* (2009). The mentioned variables progress in time t [h] and the (bioreactor) temperature T [K] influence on maximum specific cell growth rate μ_m [h^{-1}] are:

$$\{dX/dt, \, dS/dt, \, dCO_2/dt \text{ and } dP/dt\} = f(X, S, CO_2, \mu_m(T)) \tag{1}$$

The non-isothermal kinetic model is constituted by the previous mass balances and the energy balance in the reactor and its cooling jacket (Eq. 2), coupled by means of Arrhenius' equation (Aballay *et al.*, 2008). Assumptions are: heat transfers by radiation and conduction and heat losses due to CO_2 evolution, water evaporation and ethanol and flavor losses are negligible. Water properties and the bulk pH were considered constant.

$$\frac{d(\rho_r \cdot V_r \cdot Cp_r \cdot T)}{dt} = Y_{H/CO_2} \cdot V_r \cdot \frac{dCO_2}{dt} - F_w \cdot Cp_w \cdot (T_{w,out} - T_{w,in}) \tag{2}$$

V_r [m^3] is the volume. $Y_{H/CO2}$ [W·h produced/kg CO_2 released] is the energy due to the carbon dioxide released by the bio-reaction, and the second term in the right side represents the exchanged heat between the fermenting mass and the cooling jacket [W]. F_w is the cooling water flow rate [kg h^{-1}] and C_{pw} is specific heat of water [W·h kg^{-1} K^{-1}]; the cooling water temperature [K] in the jacket inlet is $T_{w,in}$ and in the jacket outlet is $T_{w,out}$ (see details in Aballay *et al.*, 2008). ρ_r [kg m^{-3}] and C_{pr} [W·h kg^{-1} K^{-1}] are density and specific heat of the fermenting mass, which were incorporated to the model as variables to achieve more accuracy in process state prediction, approximately a 10%.

3. Dynamic model performance analysis

For this purpose, steps changes on most probable manipulated variable in practice, such as the cooling water flow rate, were introduced. A decrease or shut-down in the cooling water flow can affect the most critical variable for wine quality, the temperature. Model performance was evaluated considering responses of parameter μ_m and, the T and CO_2 evolution to negative steps, -33% and -70% of nominal F_w (Fig. 1). The changes produce a positive increase on: maximum growth rate (up to +20%), (a), and maximum temperature (up to +25%), (b). Moreover, an early positive increase in carbon dioxide released (+14%) at t = 88 hours can be observed, (c). Therefore, the dynamic model performance is satisfactory.
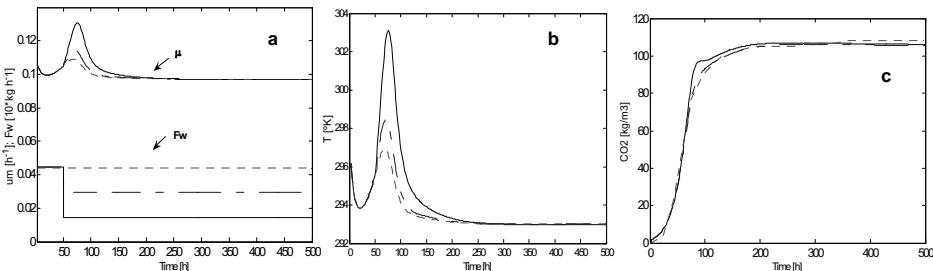
Figure 1. Model responses to steps on the cooling water flow rate. $F_{w,nominal}$ = 0.448 kg h^{-1} (- - -); $F_{w,}$ = 0.30 kg h^{-1} (— - —); $F_{w,}$ = 0.143 kg h^{-1} (—). (a) Maximum specific growth rate and water flow rate. (b) Bioreactor temperature. (c) Carbon dioxide released.

Most probable changes on other mass balances parameters in the model are shown in a wide performance analysis (Scaglia *et al.*, 2009). Regarding to the main energy balance parameters, they were varied in ±20% and the corresponding responses on the state process variables have been obtained with satisfactory results (not shown).

4. Control problem

The proposed control on temperature is based on the manipulation of cooling water flow rate through a dynamic inverse controller. Previously knowing the process model and the desired temperature profile ($T_{ref}(t)$), this control strategy allows to generate the corresponding control actions in such a way that the tracking error ($e(t)$) tend to zero. The controller has been performed in discrete-time, so that the current disturbance effect at the output is computed as the difference between the current measured value of the output and the predicted one.

The values of $T(t)$ at discrete time $t = n\,T0$, where $T0$ is the sampling period and $n \in \{0,1,2,\dots\}$, will be denoted as $T(n)$. From Eq. 2, the value of $T(n+1)$ is,

$$T(n+1) = T(n) + \int_{nT0}^{(n+1)T0} \frac{1}{\rho_r V_r Cp_r}\left[Y_{H/CO_2}\, V_r \frac{dCO_2}{dt} - F_w Cp_w \left(T_{w,out} - T_{w,in}\right) \right] dt \qquad (3)$$

where, F_w remains constant through the interval $nT0 \leq t < (n+1)T0$. Using the Euler approximation to calculate the integral, we have:

$$T(n+1) \approx T(n) + \frac{T0}{\rho_r V_r Cp_r}\left[Y_{H/CO_2}\, V_r \frac{dCO_2(n)}{dt} - F_w(n)Cp_w \left(T_{w,out}(n) - T_{w,in}(n)\right) \right] \qquad (4)$$

$$F_w(n) = \left\{ \left[T_{ref}(n+1) - k_T\left(T_{ref}(n) - T(n)\right) - T(n)\right]\frac{\rho_r V_r Cp_r}{T0} - Y_{H/CO_2}\, V_r \frac{dCO_2}{dt} \right\} \Big/$$
$$\dots\dots - \left[Cp_w\left(T_{w,out} - T_{w,in}\right) \right] \qquad (5)$$

Equation 5 has only the error due to the integral approximation, whose maximum value is A (Eq. 6) (Boyce and DiPrima, 2005), where clearly, if the sampling time is small, also the error will be small. Replacing 5 in 4 we obtain 6:

$$A = \max_t\left(\frac{1}{2}\frac{d^2T}{dt^2}T0^2 \right) \;\; ; \;\; \underbrace{T_{ref}(n+1) - T(n+1)}_{e(n+1)} - k_T\underbrace{\left(T_{ref}(n) - T(n)\right)}_{e(n)} = 0 \qquad (6)$$

In Eq. 6, it can be seen that if $0 < k_T < 1$, then $e(n) \rightarrow 0$ when $n \rightarrow \infty$, where $e(n)$ is the tracking error in discrete form. Besides, since in each sampling period the process variables are measured, there are no accumulative errors. Nevertheless, a term ($E(n)$), that takes into account the approach error of the integral and the modeling error, is included in Eq. 4 (See Eq. 7). Here, the terms with cup (ˆ) represent the estimated values of the actual system parameters. Finally, considering $E(n)$ the control action is expressed by Eq. 8.

$$T(n+1) = T(n) + \frac{T0}{\widehat{\rho}_r \widehat{V}_r \widehat{Cp}_r}\left[\widehat{Y}_{H/CO_2}\, \widehat{V}_r \frac{dCO_2}{dt} - F_w(n)\widehat{Cp}_w\left(T_{w,out} - T_{w,in}\right) \right] + E(n) \qquad (7)$$

$$F_w(n) = \left\{ \left[T_{ref}(n+1) - k_T \left(T_{ref}(n) - T(n) \right) - T(n) - E(n) \right] \frac{\widehat{\rho}_r \widehat{V}_r \widehat{C}p_r}{T0} - \widehat{Y}_{H/CO_2} V_r \frac{dCO_2}{dt} \right\} \Bigg/ \cdots\cdots - \left[\widehat{C}p_w \left(T_{w,out} - T_{w,in} \right) \right] \tag{8}$$

Figures 2a and 2b show the temperature change and the control action respectively, when the latter is calculated with Eq. 5 for a small sampling period. It can be seen how the system temperature tracks closely the reference trajectory. Also, it is observed that such tracking is achieved with a finite control action (Fig. 2b). On the other hand, errors in model parameters significantly influence the process states estimation, carried out by the multi-parametric and good performance model (Scaglia *et al.*, 2009; Aballay *et al.*, 2008), and consequently on the control law. Figure 2c shows the bioreactor temperature change when some model parameters contain error. In this case, the control signal was calculated with Eq. 8, but with $E(n) = 0$. If Figs. 2a and 2c are compared it is observed that the parameters errors produce a considerable deviation between the bioreactor and reference temperature profiles (Fig. 2c). The inclusion of term $E(n) \neq 0$ in the proposed control signal (Eq. 8) improves significantly the temperature tracking (Fig. 3a). Besides, the good performance of the system can be seen in Fig. 3b that shows the comparison with one of the literature data. X profile for the T_{ref} tracking control shows a satisfactory reduction in the cellular death (from 150 h) with respect to experimental fermentation FER1 in Aballay *et al.* (2008). In this latter case without a T_{ref} profile tracking, cellular death achieved is total.
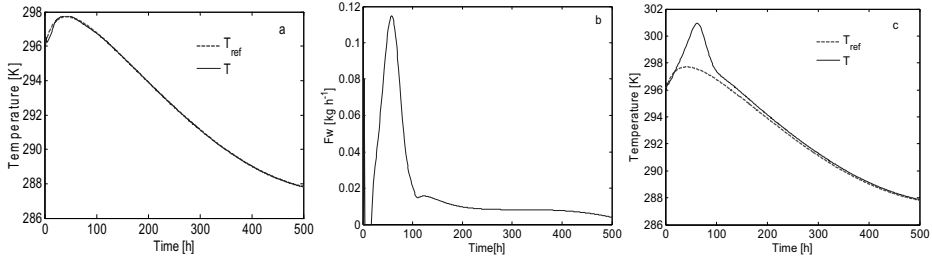
Figure 2. Temperature tracking. (a) Reference and bioreactor temperature profiles. (b) Control action on cooling water flow rate. (c) Reference and bioreactor temperature profiles when there are errors in the model parameters.
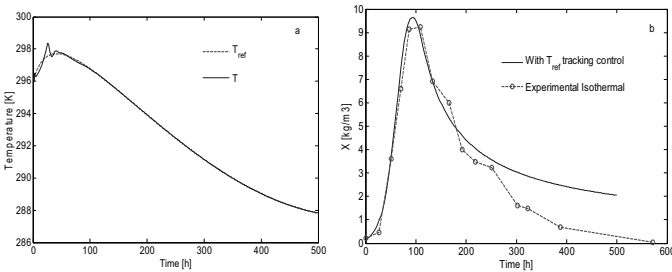
Figure 3. (a) Reference and bioreactor temperature profiles when the modelling errors are considered. (b) Viable cells profile: experimental isothermal fermentation and with the proposed control scheme for the reference temperature profile.

As can be seen from Figs. 2 and 3, the main control objective is achieved using a finite control signal, in spite of the model can have modeling errors.

5. Conclusions

An advanced temperature tracking control for high quality wines have been presented. The control system is based on an improved non-isothermal model that predicts closely the main fermentation state variables and allows a strict process control, when non-constant temperature profiles must be tracked to achieve distinctive fine wines. The proposed control manipulates the cooling water flow rate through a dynamic inverse controller. The system shows a good performance tracking narrowly a predefined reference trajectory with a finite control signal, even though the model can present errors in its parameters. The system was satisfactorily tested with literature experimental data. In next contributions, the system will be extended and improved to tracks other important process variables such as density and cell concentration.

References

P.M. Aballay, G.J.E. Scaglia, M.D. Vallejo, and O.A. Ortiz, 2008, Non isothermal phenomenological model of an enological fermentation: modelling and performance analysis, 10th International Chemical and Biological Engineering Conference CHEMPOR 2008, Braga, Portugal, 4-6 September.

G. Beltran, M. Novo, J.M. Guillamón, A. Mas, and N. Rozès, 2008, Effect of fermentation temperature and culture media on the yeast lipid composition and wine volatile compounds, International Journal of Food Microbiology, 121, 169–177.

R. Boulton, V.L. Singleton, L. Bisson, and R. Kunkee, 1996, Yeast and biochemistry of wine fermentation, Principles and practices of winemaking, Chapmann and Hall, New York.

W.E. Boyce, and R. C. DiPrima, 2005, Elementary Differential Equations and Boundary Value Problems, 8th Edition, John Wiley & Sons Limited, Australia.

A.C. Cramer, S. Vlassides, and D.E. Block, 2002, Kinetic model for nitrogen-limited wine fermentation, Biotechnol. Bioeng., 77, 49-60.

M. Esti, and P. Tamborra, 2006, Influence of winemaking techniques on aroma precursors, Analytica Chimica Acta, 563, 173–179.

M.A. Henson, 2003, Dynamic modeling and control of yeast cell populations in continuous biochemical reactors, Computers & Chemical Engineering, 27, 8-9, 1185-1199.

A.M. Molina, J.H. Swiegers, C. Varela, I.S. Pretorius, and E. Agosin, 2007, Influence of wine fermentation temperature on the synthesis of yeast-derived volatile aroma compounds, Appl. Microbiol. Biotechnol., 77, 675–687.

O.A. Ortiz, P.M. Aballay, and M.D. Vallejo, 2006, Modelling of the Killer Yeasts Growth in an Enological Fermentation by means of a Hybrid Model, XXII Interamerican Congress of Chemical Engineering, A. 13b-224, 451-452, Buenos Aires, Argentina, 1-4 October.

I. Romero Cascales, 2008, Extracción de compuestos fenólicos de la uva al vino. Papel de las enzimas de maceración, Tesis Doctoral (PhD Thesis), Universidad de Murcia, Spain.

E. Rosenfeld, B. Beauvoit, B. Blondin, and J. Salmon, 2003, Oxygen consumption by anaerobic *Saccharomyces cerevisiae* under enological conditions: effect on fermentation kinetics, Appl. Environ. Microbiol., 69, 113-121.

G.J.E. Scaglia, P.M. Aballay, C.A. Mengual, M.D. Vallejo, and O.A. Ortiz, 2009, Improved phenomenological model for an isothermal winemaking fermentation, Food Control, doi: 10.1016/j.foodcont.2008.12.012

M.J. Torija, G. Beltran, M. Novo, M. Poblet, J.M. Guillamón, A. Mas, and N. Rozès, 2003, Effects of fermentation temperature and Saccharomyces species on the cell fatty acid composition and presence of volatile compounds in wine, Int. J. Food Microbiol., 85, 127–136.

M.D. Vallejo, P.M. Aballay, M.E. Toro, F. Vazquez, G.I. Suarez, and O.A. Ortiz, 2005, Hybrid Modeling and Neural Prediction of the Wild Killer Yeast Fermentation Performance in a Winemaking Process, 2nd Mercosur Congress on Chemical Engineering and 4th Mercosur Congress on Process Systems Engineering, Paper code 230, 1-10, Rio de Janeiro, Brazil, 14-18 August.

1395

MPC of a Four-Stage Grape Juice Evaporator Based on an Adaptive Neural Model

Graciela I. Suárez, Gustavo J.E. Scaglia, Pablo M. Aballay and Oscar A. Ortiz

Instituto de Ingeniería Química, Universidad Nacional de San Juan, Av. Lib. San Martín Oeste 1109, J5400ARL San Juan, Argentina

Abstract

The four-stage evaporator is the core of the process in the manufacture of concentrated grape juice. The dynamic features of this process are very complex due to inputs/outputs constraints, time delays, loop interactions and the persistent unmeasured disturbances that affect it. This paper addresses the non-linear control of an industrial-scale multiple-effect evaporator by means of a model predictive control (MPC) based on a neural network model (NNM). This strategy allows it on one hand, overcome the classical control systems limitations and, by the other that the neural controller continue learning during the plant operation. The achieved results allow us to conclude that the developed neural model predictive control is adequate to control effectively a four-effect evaporator to concentrate grape juice with different characteristics; its application at industrial scale is possible in the future.

Keywords: Adaptive model-based predictive control; artificial neural networks; grape juice concentration process

1. Introduction

The multi-step evaporation (M-SE) is the most important unit operation used in the food industry to concentrate juices of grapes and others fruits. Even when the main objective of this process is to produce a concentrated product, this should also possess certain organoleptic properties that are critical with respect to its quality and acceptance grade by the customers.

Product requirements and the complex characteristics of the process such as non-linear behavior, input and output constraints, time delays and loop interactions justify the use of an advanced control system. In this sense, the control of industrial evaporation processes with such constrains remains as a challenge for researchers.

Diverse control strategies have been intended for multi-stage evaporators. Among others the following works can be mentioned: conventional multi-loop linear feedback control, differential geometric control, generic model control GMC, multivariable generalized predictive control GPC, input-output linearization, differential geometric control using input-output linearization. Some authors have applied with success the differential geometric control to evaporation systems in the alumina industry (Kam and Tadé, 1999; Kam and Tadé, 2000). On the other hand, the model based predictive control (MPC) is today widely used in the industry and recognized as a powerful tool for high performance control of constrained multivariable processes. The main features of some of these industrial applications are available in Camacho and Bordons (1999).

In this work, a model predictive control (MPC) based on a neural network model (NNM) is presented. The neural network model is trained with data generated by simulation of a previously developed phenomenological model and with a portion of

plant data. The phenomenological model validation as well as the plant data used for NN training corresponds to a 1370 kg/h-production four-effect evaporator. The simulation data used for NN training include a range of most likely plant operating conditions introduced by means of pseudo-stochastic disturbances in the form of white noise on the manipulated variables; they are: feed flow rate and/or the heating steam temperature and, the range of variation is about ±10%. Afterwards, the NN model is validated with the other portion of the input-output plant data. This adaptive strategy is adopted, in order to obtain a more accurate process description without introducing disturbances in the real plant, which supposes a high cost, to obtain relevant data sets for training the NN.

In spite of a phenomenological model is available it is preferred the use of a NN model trained with data from simulation. The lack of sufficient amount of plant data does not permit an adequate training of the NN, however this approach has the advantage that at the beginning, the NN trained by means of simulation, at least imitate the performance of the phenomenological model. Therefore, when the NN model is coupled with the real plant in a MPC it can continue learning from the process and improve its performance through time. The first principles model is not used in the control strategy proposed because it has the inherent errors to the modelling of complex processes and, it does not have the capacity to adapt to the changing operation conditions of this kind of processes (Psichogios and Ungar, 1991; Hussain et al., 2003). NNs of different sizes are trained and those ones with the best performance with respect to the first principles model are selected.

The paper is organized as follow. In Section 2, a short description of a four-effect evaporator is presented, which is described by a set of non-linear ODEs. Section 3 presents the proposed control scheme, where mathematical and neural models are used to compute the control signal with the objective of considering the model variations due to changes in composition, organoleptic and physical properties of the grape-juice to be processed. Section 4 presents the discussion of the main obtained results, analyzes the contribution of the work and outlines some topics to be addressed in future works.

2. Process Description

Figure 1a shows the input and output streams in a vertical generic effect evaporator with long tubes. The solution to be concentrated circulates inside the tubes, while the steam, used to heat the solution, circulates inside the shell around the tubes.

The evaporator operates in co current. The solution to be concentrated and the steam are fed to the first effect by the bottom and by the upper section of the shell, respectively. Later on, the concentrated solution from the first effect is pumped to the bottom of the second effect, and so on until the fourth effect. On the other hand, the vapor from each effect serves as heater in the next one. Finally, the solution leaving the fourth effect attains the desired concentration.

Each effect has a baffle in the upper section that serves as a drops splitter for the solution dragged by the vapor. The vapor from the fourth effect is sent to a condenser and leaves the process as a liquid. The concentrated solution coming from the fourth effect is sent to a storage tank.

Stefanov and Hoo (2003, 2004) have developed a rigorous model with distributed parameters based on partial differential equations for a falling-film evaporator, in which the open-loop stability of the model to disturbances is verified. On the other hand, various methods have been proposed in order to obtain reduced-order models to solve such problems (Christofides, 1998; El-Farra, Armaou and Christofides, 2003; Armaou

and Christofides, 2002; Hoo and Zheng, 2001; Zheng and Hoo, 2002). However, there is not a general framework yet, which assure an effective implementation of a control strategy in a multiple effect evaporator.

In practice, due to a lack of measurements to characterize the distributed nature of the process and actuators to implement such a solution, the control of systems represented by partial differential equation (PDE) in the grape juice evaporator, is carried out neglecting the spatial variation of parameters and applying lumped systems methods. However, a distributed parameters model must be developed in order to be used as a real plant to test advance control strategies by simulation.

In this work, it is used the mathematical model of the evaporator developed by Ortiz *et al.* (2006), which is constituted by mass and energy balances in each effect. The assumptions are: the main variables in the gas phase have a very fast dynamical behavior, therefore the corresponding energy and mass balances are not considered (Newell and Lee, 1989). Heat losses to surroundings are neglected and the flow regime inside each effect is considered as completely mixed (Newell and Fisher, 1972). Figure 1b shows real and simulated responses of the system, where mathematical model tracks accurately the real system. The model parameters were adjusted for one variety of grape juice. The advantage of working with on-line adaptive neural networks is that the network learns the model with new organoleptic properties and adapts to the characteristics of each grape juice variety, according to the variety of grape used.
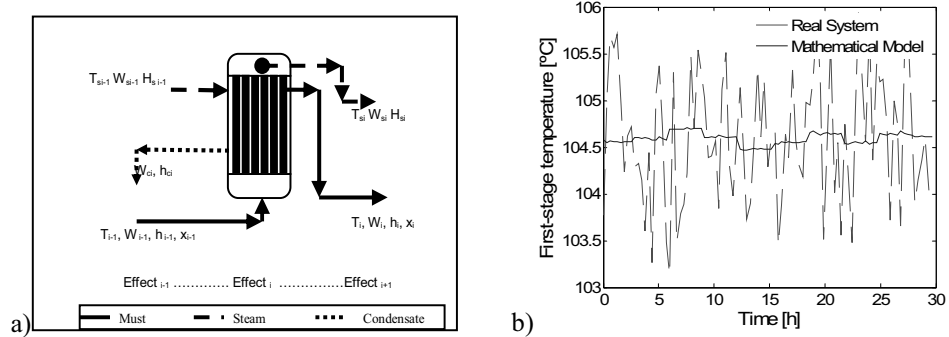
Figure 1. (a) Effect i in the four-stage evaporator flowsheet, i = 1,…, 4. (b) Real system and mathematical model responses.

3. Model-based predictive control (MPC)

The MPC is composed by two main components: the plant model and the controller – optimizer. Figure 2a shows the scheme implemented in the process. For the MPC (Camacho and Bordons, 1999), the following variables that determine its operation are defined: prediction horizon, the penalty coefficient of the control signal, and the control horizon.

The MPC algorithm allows defining future control actions while the next function is minimized.

$$J = \sum_{N_{1k}}^{N_{2k}} \gamma_k (\hat{y}_k(t-j) - Yref_k(t+j))^2 + \sum_{1}^{Nu_k} \beta_k (u_k(t+j-1) - u_k(t+j-2))^2 \qquad (1)$$

N_{1k} and N_{2k} constitute the initial prediction horizon for each output and N_{uk} is the control horizon for each input. y is the first-stage temperature, Y_{ref} is the temperature profile to

be tracked by the system and u is the heating steam temperature that enters to evaporator.

The success of Model Predictive Control (MPC) control performance is highly dependent on the accuracy of the open-loop predictions, which in turn depends on the accuracy of the process models. It is possible for the predicted trajectory to differ, perhaps considerably, from the actual trajectory followed by the plant being controlled (Berber, 1995). The difference between the plant and the model is known as plant-model mismatch, which can cause the control performance to be sluggish, overly conservative or, in the worst-case scenario, unstable. On the other hand, closed-loop performance may be poor if the plant differs significantly from the model. In the case of the evaporator, grape juice physicochemical parameters change with the varietal type to be processed. In Fig. 2b shows the control system response, when the phenomenological model is used in the MPC scheme, for a step change introduced in the reference, when the used model accurately describes the plant and when there are errors in the model. As can be seen, when the model precisely describes the plant to be controlled, the output signal tracks the reference signal in accurate way, whereas when there are errors in the model, the plant increases the difference between reference and system response signals, respectively. A neural network is used as model to solve this problem. The plant model identification was carried out using a radial basis function neural network (RBF-NN), due to its complexity. The training of the network was carried out by means of the Neural Network Toolbox of Matlab^{TM}. For the training process, the Levenberg - Marquardt Learning Law was used with 2000 pair of input – output values obtained by simulation. The simulator was excited with a pseudo-stochastic multi-binary signal (PRMS) over the manipulated variables (heating steam temperature), to obtain signals which produce an enough wide variation in the system operation conditions.
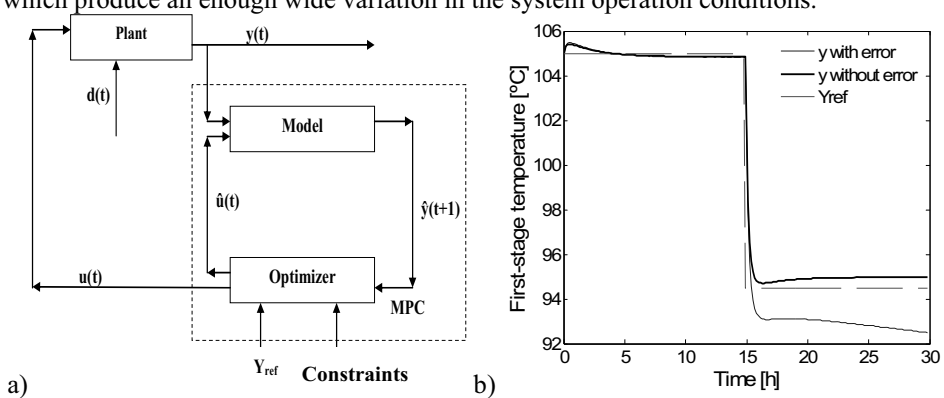
Figure 2. (a) MPC scheme. (b) System response with and without modelling errors when a step change occurs in the reference output.

A data set with 1000 pairs of values was used for training and other 1000 pairs were used in the validation stage. The sum of the quadratic errors between the NN model output and the simulation results from the validation data set have been considered in order to select the most suitable NN. The network with the better performance has 15 neurons in the hidden layer and a linear function in its output layer. The obtained neural model presents a maximum error less than 4% with respect to validation data. Then, this model is used for implementing the predictive control strategy so that the NN learns the plant behavior. Figure 3a shows the system response when the NN is used as the plant model. It can be seen that, when there are errors between the model used to train the NN and the plant to be controlled, the network can learn that difference and improve the

system performance. Moreover, it can be seen how the error is decreasing while the NN learns the system behavior and when a step change in the reference occurs, the system shows a better behavior than the observed in Fig. 2b, where the used model has errors in its parameters. In Fig. 3b it can be seen that the control objective is achieved with a bounded control signal.
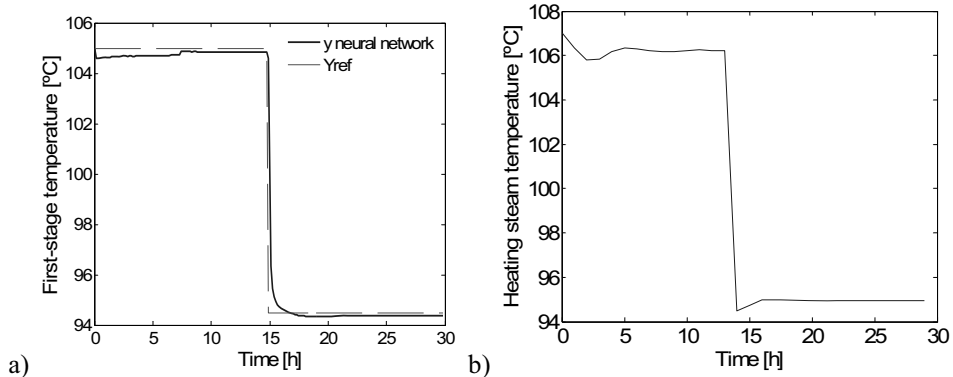
Figure 3. (a) System response when the model is a NN. (b) Control action.

4. Conclusions

In this work, a model-based predictive controller for an industrial-scale four-stage grape juice evaporator is presented, considering two alternative scenarios. First, the MPC was implemented with a phenomenological model with its parameters adjusted for a specific grape juice. Second, a NN model, obtained with data from simulation of a phenomenological model validated with industrial data, was used for the MPC. When there are no modelling errors both control alternatives show a suitable performance. In case of the model has errors, i.e., due to organoleptic and physicochemical properties variations in the grape juice, it is observed that by using a NN model, the MPC controller has a better system performance, since the NN can learn and correct the difference between the phenomenological model and the process to be controlled. The use of a MPC based on an on-line adaptive NN model allows that the controller responds adequately when the process is fed with grape juices that contain different organoleptic properties. These properties vary with the grape variety to be used. The development of an advanced control algorithm using a NN which takes variables coupling (first-stage temperature and final product concentration) into account as process model, will be addressed in a next work.

References

A. Armaou, and P.D. Christofides, 2002, Dynamic Optimization of Dissipative PDE Systems Using Nonlinear Order Reduction. Chemical Engineering Science 57, 24, 5083-5114.

R. Berber, 1995, Methods of Model Based Process Control, in NATO ASI Series. Series E, Applied Sciences, 293, Kluwer Academic Publishers, Dordrecht (The Netherlands).

E.F. Camacho, and C. Bordons, 1999, Model Predictive Control, Springer-Verlag.

P.D. Christofides, 1998, Robust Control of Parabolic PDE Systems, Chemical Engineering Science, 53, 16, 2949-2965.

N.H. El-Farra, A. Armaou, and P.D. Christofides, 2003, Analysis and Control of Parabolic PDE Systems with Input Constraints, Automatica, 39, 4, 715-725.

K.A. Hoo, and D. Zheng, 2001, Low-Order Control Relevant Model for a Class of Distributed Parameter Systems, Chemical Engineering Science, 56, 23, 6683-6710.

M.A. Hussain, 2003, in Cornelius T. Leondes (Ed.), Neural Network Techniques and Application in Chemical Process Control System, 5, V327–V360, CRC Press.

K.M. Kam, and M.O. Tadé, 1999, Case studies on the modelling and control of evaporation systems, XIX Interamerican Congress of Chemical Engineering COBEQ.

K.M. Kam, and M.O. Tadé, 2000, Simulated Nonlinear Control Studies of Five Effect Evaporator Models, Computers and Chemical Engineering, 23, 1795 - 1810.

R.B. Newell, and P.L. Lee, 1989, Applied process control: A case study, Prentice Hall.

R.B. Newell, and D.G. Fisher, 1972, Model development, reduction and experimental evaluation for an evaporator, Eng. Chemical Process Res. Develop., 11, 2, 213-221.

O.A. Ortiz, G. I. Suárez, and C. A. Mengual, 2006, Adaptive Neural Model Predictive Control for the Grape Juice Concentration Process, in Proceedings of XXII Interamerican Chemical Engineering Congress and V Argentinian Congress of Chemical Engineering. Innovation and Management for Sustainable Development, Buenos Aires, Argentina, 1-4 October.

D.C. Psichogios, and L.H. Ungar, 1991, Direct and indirect model based control using artificial neural networks, Ind. Eng. Chem. Res., 30, 12, 2564-2573.

Z.I. Stefanov, and K.A. Hoo, 2003, A Distributed-Parameter Model of Black Liquor Falling Film Evaporators. Part I. Modeling of Single Plate, Ind. Eng. Chem. Res., 42, 1925-1937.

Z.I. Stefanov, and K.A. Hoo, 2004, Distributed Parameter Model of Black Liquor Falling-Film Evaporators. 2. Modeling of a Multiple-Effect Evaporator Plant, Ind. Eng. Chem. Res., 43, 8117-8132.

D. Zheng, and K.A. Hoo, 2002, Low-Order Model Identification for Implementable Control Solutions of Distributed Parameter Systems, Computers and Chemical Engineering, 26, 7-8, 1049-1076.

1401

On l_1-Predictive Control of Mixed-Logical Dynamical Systems

David L. de Souza Júnior[a], Luís C. Oliveira-Lopes[a]

[a]*FEQUI/UFU – School of Chemical Engineering, Federal University of Uberlandia, Av. João Naves de Ávila, 2121, Santa Mônica, CEP 38408-902, Uberlândia – MG , Brazil*

Abstract

Constant technological innovations and environmental and economical requirements have created new challenges in industrial process control. The hybrid system predictive control is a representative example of great interest in this modern scenario. Hybrid behavior is found by continuous operation dynamics inserted with discrete event changes that might occur by internal and external actions of the system. Internal actions that generate hybrid behavior can be faults, changes between operation modes and perturbations. External actions can be represented by logical and decisory aspects and safety constraints. The chemical process operation often deals with controlled transitions among operation modes due to changes in the raw materials, energy sources, product specification and market demand. In this paper, a hybrid systems representation of linear systems known as mixed logical dynamical models (MLD) is investigated. The paper addresses the MLD l_1-Model Predictive Control (MPC) problem. Structurally, the MLD model control problem is described through a linear mixed integer optimization problem (MILP), which is computationally demanding, but they can be satisfactorily applied to systems of certain speed and dimension. The paper presents a study on the characteristics of the controller. A system of interest is used to illustrate the study. The results show that the application of MLD in the MPC control framework can have a positive impact specially when there is the inclusion of logical decisions aspects and operational knowledge in the MPC problem.

Keywords: Hybrid systems, MPC, MLD.

1. Introduction to Hybrid Systems

The term *hybrid* refers basically to a mixture of two or more methods or objects. The expression *hybrid system* was originally proposed by Witsenhausen (1966) to describe a combination between continuous dynamical systems and discrete event systems. Hybrid Systems come from the coupling between digital controllers, computers and subsystems built as finite-state machines, and plants modeled by partial or ordinary differential equations, or difference equations (Lazar, 2006).

Hybrid systems can be composed by several operating modes, endowed with a proper equations set, therefore they are capable of showing various dynamical behaviors simultaneously. Amongst these behaviours are continuous and discrete-time dynamics, jump discontinuity and transitions driven by logical commands. These systems are capable to select the activation of a specific mode when a particular event occurs. Figure 1 shows this fact. In a transition among operating modes, discontinuities may happen.
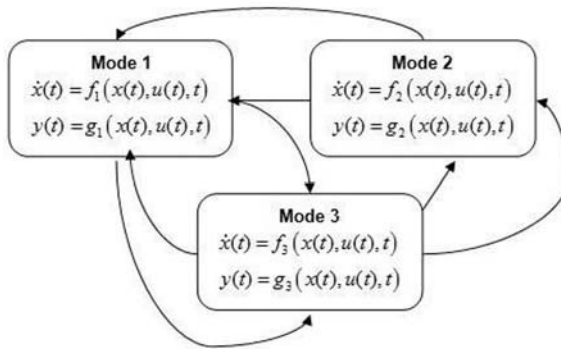
Figure 1 – Hybrid system sketch composed by three operational modes. The transitions are caused by events.

An event can be classified according to its origin. A *state event* occurs when variables reach a specific value. A *time event* occurs by elapsing a definite time period. And an *input event* occurs when external inputs influence the system (Lazar, 2006).

Although hybrid systems are capable to retain an appreciable variety of dynamic behaviors, a hybrid system general formulation introduces high levels of complexity for analysis purposes, application of techniques to design of controllers, simulation and validation. This fact is responsible for the form as the research in hybrid systems is led. Instead of a general formalism, specific classes of systems are adopted to represent, with relative simplicity and fidelity, the process of interest. Distinguished classes of hybrid systems are: *Piecewise affine systems*, PWA (Sontag, 1981, Johansson and Rantzer, 1998, Ferrari-Trecate et al., 2002); *Mixed-logical dynamical systems*, MLD (Bemporad and Morari, 1999, Bemporad, 2004); *Linear complementarity systems*, LC (van der Schaft and Schumacher, 1998; Heemels et al., 2000); and *Max-min-plus-scaling* systems (MMPS) (De Schutter and van den Boom, 2001).

1.1. Mixed-logical dynamical systems

Mixed-logical dynamical systems are described by equations set 1:

$$x(k+1) = A_k x(k) + B_{1k} u(k) + B_{2k} \delta(k) + B_{3k} z(k) \tag{1.a}$$

$$y(k) = C_k x(k) + D_{1k} u(k) + D_{2k} \delta(k) + D_{3k} z(k) \tag{1.b}$$

$$E_{2k} \delta(k) + E_{3k} z(k) \geq E_{1k} u(k) + E_{4k} x(k) + E_{5k} \tag{1.c}$$

where $k \in \mathbb{Z}$; A_k, $B_{1k,2k,3k}$, C_k, $D_{1k,2k,3k}$, $E_{1k,2k,3k,4k,5k}$ are matrixes with appropriate dimensions. And also, the following domains are defined by the set 2:

$$x = \begin{bmatrix} x_c & x_l \end{bmatrix}^T, x_c \in \mathbb{R}^{n_c}, x_l \in \{0,1\}^{n_l}, n = n_c + n_l$$

$$y = \begin{bmatrix} y_c & y_l \end{bmatrix}^T, y_c \in \mathbb{R}^{p_c}, y_l \in \{0,1\}^{p_l}, p = p_c + p_l \tag{2}$$

$$u = \begin{bmatrix} u_c & u_l \end{bmatrix}^T, u_c \in \mathbb{R}^{m_c}, u_l \in \{0,1\}^{m_l}, m = m_c + m_l$$

Vector **x** denotes the states, which are composed with continuous states ("c" index), and logical binary states, 0-1 ("l" index). The vector **y** represents the model output data. The input vector **u** encloses continuous commands ("c" index) and binary commands, 0-1 ("l" index). The MLD system is said to be well posed if $\delta(k)$ and $z(k)$ are uniquely defined by inequality 1.c, in its domain, once **x**(k) and **u**(k) are specified. Thus, from

equation 1.a and 1.b, **x**(k+1) e **y**(k) are exclusively functions of **x**(k) and **u**(k). An interesting advantage is, in certain conditions, MLD systems can be equivalent to the other model classes in the section 1 (Heemels et al., 2001). This fact means that systems modeled by this kind of description can be converted into MLD systems that possess well definite techniques that favor its use for optimal and predictive control purposes.

In the MLD formulation, logical propositions can be rewritten in the format of linear inequalities by means of the transformation of Boolean variables in binary integer variables. Previous operational knowledge of a process can also be added systematically under the form of inequalities. Together with the equations of type 1.a and 1.b, which dictate the dynamics of the process, these linear inequalities form the set of constraints of an optimization problem which is base for optimal control problem synthesis (Bemporad e Morari, 1999; Krilavicius, 2006).

2. MLD Systems Optimal Control Aspects

Optimal control target is to make possible the synthesis of control law which deals with certain purposes. Such purposes are congregated in a performance index. It is desirable that this index is minimized subject to the process constraints. It is possible to penalize efforts of control (energy) and input/output reference deviation.

Control optimal synthesis yields an optimization problem. The resulting optimization problem for a MLD system can have difficult solution; therefore it involves mixed-integer programming (MILP, MIQP, and MINLP). It is also complicated the performance index selection and this issue demands an iterative procedure for its adequacy to the problem.

2.1. Definition of MLD system predictive control

Given to an initial state x_0 and a final instant T, it is desirable to find, if possible, a sequence of control **u**, defined as: $u \equiv \{u\ (0),\ u\ (1),...,\ u\ (T\text{-}1)\}$, that leads the system of the state x_0 to x_f, minimizing the following performance index:

$$J\left(u_0^{T-1}, x_0\right) = \sum_{k=0}^{T-1} \left\|\Delta u(k) - \Delta u_f\right\|_{Q_1}^p + \left\|\delta\left(k, x_0, u_0^k\right) - \delta_f\right\|_{Q_2}^p + \left\|z\left(k, x_0, u_0^k\right) - z_f\right\|_{Q_3}^p$$

$$+ \left\|x\left(k, x_0, u_0^{k-1}\right) - x_f\right\|_{Q_4}^p + \left\|y\left(k, x_0, u_0^{k-1}\right) - y_f\right\|_{Q_5}^p \tag{3}$$

Subject to the MLD system and its constraints, and also to a terminal/stability condition, such as, for instance:

$$x\left(T, x_0, u_0^{T-1}\right) = x_f \tag{4}$$

$\|x\|_{Q_i}^p$ represents the l_p-norm application under vector **x**. $Q_i = Q_i' \geq 0$, i = 1,2,..., 5,

weight matrixes and x_f, u_f, δ_f, z_f, y_f represent a system equilibrium point that satisfies the equations set 1. From l_p-norm definition, l_1-norm is established below, where $x^{(i)}$ represents the ith entry of a given vector **x**:

$$\| x \|^p = \left(\sum_{i=1}^{n} \left| x^{(i)} \right|^p \right)^{1/p} = \sum_{i=1}^{n} \left| x^{(i)} \right| \tag{5}$$

The biggest advantage of the use of a linear performance index is that the optimization problem can be written as a linear programming problem (LP). Solving LP problems is less demanding, in a computational effort point of view, than the quadratic

programming (QP) for same size and complexity systems (Rao and Rawlings, 2000). The same can be said for problems of mixed-integer programming (MIP). A mixed-integer linear programming (MILP) problem is much less demanding than mixed-integer quadratic programming (MIQP) problems. While for simple LP and QP problems this might not be that relevant, for MILP and MIQP problems this difference is somewhat more important.

2.2. MLD optimal control as MILP

When l_1-norm is applied, the equation 3 gives the following performance index:

$$J\left(u_0^{T-1}, x_0\right) = \sum_{k=0}^{N-1} \sum_i Q_1 \left|\Delta u^{(i)}(k) - \Delta u^{(i)}{}_f\right| + \sum_i Q_2 \left|\delta^{(i)}(k) - \delta^{(i)}{}_f\right|$$
$$+ \sum_i Q_3 \left|z^{(i)}(k) - z^{(i)}{}_f\right| + \sum_i Q_4 \left|x^{(i)}(k) - x^{(i)}{}_f\right| + \sum_i Q_5 \left|y^{(i)}(k) - y^{(i)}{}_f\right| \tag{6}$$

where index "i" is the ith entry of the respective vector. It is necessary, therefore, to write the optimal control problem, enunciated in section 2.1, in the MILP problem format. Thus, auxiliary project variables are adopted to transfer the modular terms to the constraints and the rewritten performance index is:

$$J\left(u_0^{T-1}, x_0\right) = \sum_{k=0}^{T-1} \left(\sum_i \alpha_i + \sum_i \beta_i + \sum_i \eta_i + \sum_i \mu_i + \sum_i \rho_i\right) \tag{7}$$

The optimization problem solved in this work may be written as given by Eq. (8).

$$\min \gamma$$

$$subject\, to: -\alpha_k \leq Q_1\left(\Delta u(k) - \Delta u_f\right) \leq \alpha_k$$
$$-\beta_k \leq Q_2\left(\delta(k) - \delta_f\right) \leq \beta_k$$
$$-\eta_k \leq Q_3\left(z(k) - z_f\right) \leq \eta_k$$
$$-\mu_k \leq Q_4\left(x(k) - x_f\right) \leq \mu_k \tag{8}$$
$$-\rho_k \leq Q_5\left(y(k) - y_f\right) \leq \rho_k$$

$$0 \leq \sum_{k=0}^{N-1} \left(\sum_i \alpha_i + \sum_i \beta_i + \sum_i \eta_i + \sum_i \mu_i + \sum_i \rho_i\right) \leq \gamma$$

$$and\, MLD\, system\, equations$$

For this predictive control problem solution, the free software Scilab 4.1.2® was used in association with LPSOLVE optimization applicative interface, (LPSOLVE 5.5 API for Scilab). LPSOLVE code congregates several linear optimization techniques, amongst them, the Branch and Bound technique, which is capable to handle with mixed-integer linear programming with easy.

It is clear that there is already software available in the literature for formulating similar problems, such as the Hybrid Toolbox (Bemporad, 2004a) that uses HYSDEL models. Nevertheless, this paper aims to use the l_1-norm out of that context, aiming to show that it is easy to formulate and viable to be implemented for small processes. For classic MPC formulation, there are important differences between LP and QP MPC

implementations, perhaps the fact that the LP problem is easier to be solved than the QP problem is not of major concern because QP algorithms are quite developed nowadays. This is not quite true for mixed integer optimization problems. Besides that, another point deserves attention, the solution of the MPC problem for different norms are indeed different, and therefore, one needs to verify which properties are mostly wanted overall. One great advantage of QP-MPC formulations is the fact that it gives linear behavior as long as the constraints are inactive. In the hybrid description, where one has a large number of constraints, the l_1-norm keeps the solution faster, and in the intersections of hyperplanes that gives optimal solutions.

3. Application and Results

For illustrating the application of such a controller, it will be used a toy example presented by Bemporad and Morari (1999):

$$x(k+1) = 0.8 \begin{bmatrix} \cos\alpha(k) & -\sin\alpha(k) \\ \sin\alpha(k) & \cos\alpha(k) \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k),$$

$$y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k),$$

$$\alpha(k) = \begin{cases} \dfrac{\pi}{3} & if \ \begin{bmatrix} 1 & 0 \end{bmatrix} x(k) \geq 0 \\[2ex] -\dfrac{\pi}{3} & if \ \begin{bmatrix} 1 & 0 \end{bmatrix} x(k) < 0 \end{cases}, \qquad (9)$$

$$x(k) \in [-10,10] \times [-10,10],$$

$$u(k) \in [-1,1]$$

This system can be interpreted as a piecewise linear time invariant system (PWLTI) that contains two operational modes. The activation of the modes depends upon a state event, defined by α(k). Following the same lines presented in Bemporad and Morari (1999), this system can be expressed by an equivalent MLD formulation. For the sake of brevity, this transformation is omitted. Figure 2 shows the dynamical simulation of the studied system.
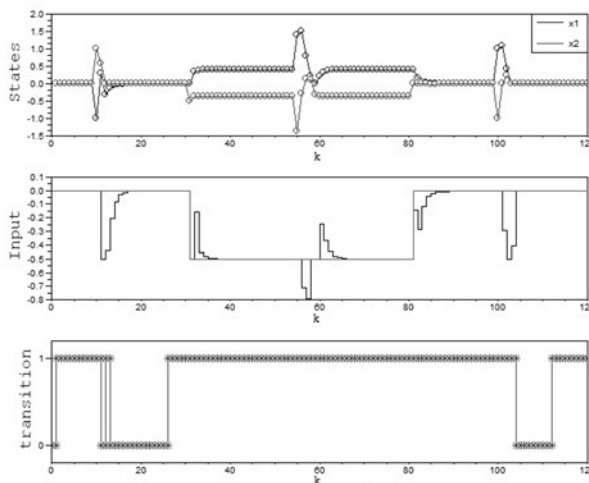
Figure 2 – Dynamical simulation: States behavior, input and its setpoint, and transition between modes.

The proposed MPC controller uses a horizon prediction H_p=3. The optimal control problem, equation set (8), presents 122 linear constraints and 53 project variables, amongst them, 6 variables are integer. The speed of control is limited in the interval: $\Delta u \in [-0,5;0,5]$. In the instants k=10, k=55 and k=100, perturbations affect the system configuring a regulatory control problem. In the instants k=30 and k=80, the setpoint is changed performing a servo control problem. The controller was also able to drive the system to its setpoint in a tracking control problem

4. Conclusions

It is known that MLD formulation allows the inclusion of logical aspects and constraints in an efficient way and that the resultant models can be used for prediction in MPC control algorithms. The selection of l_1-norm performance index makes possible the use of mixed-integer linear programming, MILP. Thus, l_1-predictive control for MLD systems can be considered a promising alternative for complex problems.

In this paper, the speed of control, Δu, is limited. This fact favors the use of l_1-norm based controllers avoiding typical problems like dead-beat and idle control. LPSOLVE 5.5 has been very efficient in MILP calculations for the simulations. Fast and stable, this free code has, basically, no limitations. The used technique for the attainment of algorithm control yields formulations that might bring problems due to the high dimensional space for optimization. However, for the current computers, this fact does not have a relevant impact for small systems and this might have an important impact when selecting or not when to use an l_1-norm or l_1-norm cost function.

References

M. Lazar, Model Predictive Control of Hybrid Systems: Stability and Robustness. Ph.D Thesis. Technische Universiteit Eindhoven, Holanda, 2006.

T. Krilavicius, Hybrid Techniques for Hybrid Systems. PhD Thesis. University of Twente, Holanda, 2003.

E. D. Sontag, Nonlinear regulation: the piecewise linear approach. IEEE Transactions on Automatic Control 26 (2), 346–357, 1981.

M. Johansson, A. Rantzer, Computation of piecewise quadratic Lyapunov functions for hybrid systems. IEEE Transactions on Automatic Control 43 (4), 555–559, 1998.

Ferrari-Trecate et al. Analysis of discrete-time piecewise affine and hybrid systems. Automatica 38 (12), 2139–2146, 2002.

A. Bemporad, M. Morari, Control of systems integrating logic, dynamics, and constraints. Automatica 35 (3), 407–427, 1999.

A. Bemporad. Efficient conversion of mixed logical dynamical systems into an equivalent piecewise affine form. IEEE Transactions on Automatic Control 49 (5), 832–838, 2004.

A. Bemporad. Hybrid Toolbox - User's Guide, availabe in http://www.dii.unisi.it/hybrid/toolbox, (2004a).

Heemels et al. Equivalence of hybrid dynamical models. Automatica 37 (7), 1085–1091, 2001.

Heemels et al. Linear complementarity systems. SIAM journal on applied mathematics 60 (4), 1234–1269, 2000.

B. de Schutter, T. J. J. Van den Boom, Model predictive control for max-plus-linear discrete event systems. Automatica 37 (7), 1049–1056, 2001.

1407

Fuzzy Model Predictive Control Applied to Piecewise Linear Systems

Thiago V. Costa[a], Lívia M. Tizzo[b], and Luís C. Oliveira-Lopes[c]

[a]*LCAP/DESQ - Department of Chemical Systems Engineering, School of Chemical Engineering, University of Campinas, Av. Albert Einstein, 500, CP 6066, CEP13083-970, Campinas-SP, Brazil*
[b,c]*FEQUI/UFU –School of Chemical Engineering, Federal University of Uberlândia, Av. João Naves de Ávila, 2121, Santa Mônica, CEP 38408-902, Uberlândia – MG, Brazil*

Abstract

This paper presents the application of a combined control strategy applied to nonlinear systems and grounded in a MPC structure using a fuzzy model description. The nonlinear models were treated as several sub-models with linear behavior, called Piecewise Linear systems (PWL). The advantages of this methodology were visualized in the temperature control of a continuous tank reactor with output multiplicity behavior. Comparisons with classical control approaches showed that the PWL MPC is an attractive and practical strategy.

Keywords: fuzzy control, MPC, piecewise linear systems (PWL)

1. Introduction

Model Predictive Control (MPC) has called a great deal of attention in the industry, and today is the controller of choice for many areas of chemical and petrochemical industry. The great advantage of the linear model based predictive control (LMBPC) approach consists in solving a convex cost function instead of a non convex one, which has no guarantees to be solved in time. However, due to the intrinsic nonlinearity of chemical industries, obtaining good first-principles model of a process is a non-trivial task. Furthermore, an input-output data based identified model has its quality severely reduced when it is moved away from its designed operating point. An efficient alternative is the utilization of fuzzy modeling that can process numerical or language information and thus have the possibility of inclusion of qualitative information in the description of the process plant to be controlled.

Literature on fuzzy models (Takagi & Sugeno, 1985) for model predictive control covers both the utilization of fuzzy black-box models (Jang, 1993; Babuska et al., 1998) and the approximation of nonlinearities characteristics using multiple linear models. Fischer et al. (1997) and Espinosa et al. (1998) investigated the utilization of an identified fuzzy model in parallel with the MPC structure which generates step response coefficients at every sampling time. Roubos et al. (1999) showed the utilization of TS models as linear models with state dependent parameters. Huang et al. (2000) reported the approximation of nonlinear models using multiple step response convolution models; Marusak (2007), in a similar but more efficient way, designed a fuzzy model predictive controller applied to a non-minimum phase reaction system.

This paper aims to design and investigate the predictive control based on fuzzy logic models. The multiple step response approach presented by Marusak (2007) is examined

and discussed; it is known that the great drawback of this formulation is the over parameterized models and its restriction to describe only asymptotically stable plants. As an alternative for this problem, a piecewise linear state space model approach is proposed. Although less intuitive than the step response models, state space models are more informative, can easily treat multiple input – multiple output (MIMO) problems and are well suited for non-stable plants as well. The strategy and the developed controllers are illustrated in the control of a highly nonlinear continuous reactor. Results and comparisons with classical techniques of process control are addressed in the study of the MPC introduced in this paper.

2. Piecewise Linear Model Predictive Control

In the basic structure of MPC, both the model and optimizer act together in order to handle future and past inputs and outputs, cost function and reference trajectory generating the future errors and dealing with constraints. In order to compute de optimal control signal, an objective function *(1)* is minimized, subject to constraints in the input, input speed, and output.

$$J = \sum_{k=H_w}^{H_p} \left\| \hat{y}(t+k) - w(t+k) \right\|_Q^2 + \sum_{k=0}^{H_c-1} \left\| \Delta\hat{u} \right\|_R^2 \qquad (1)$$

Here H_w is the process delay, H_p is the predicted horizon, H_c is the control horizon, $\hat{y}$ is the output of the plant model, w is the reference trajectory and $\Delta\hat{u}$ is the future control increment. For a control horizon $Hc>1$, a vector of inputs are returned by the optimization, but only the first term is really sent to the process. For a detailed revision of MPC, see Maciejowski (2002).

Let us assume a nonlinear process represented by *(2)*, where x is the state vector and u is the input of the process.

$$\frac{dx}{dt} = f(x,u) \quad \text{and} \quad y = g(x,u) \qquad (2)$$

a linear model could only represent the dynamic characteristics of the process in the neighborhood of a specified equilibrium point; restricting the use of the predictive controller to the boundaries of the linear model. As an alternative, the nonlinear model can be described by several linear sub-models known as piecewise linear models (PWL), which in conjunction represents the whole dynamic behavior of the nonlinear system. The PWL methodology allows the switching over the entire representation of the process (see Figure 2.1), minimizing the prediction quality losses caused by movements over the operational trajectory.
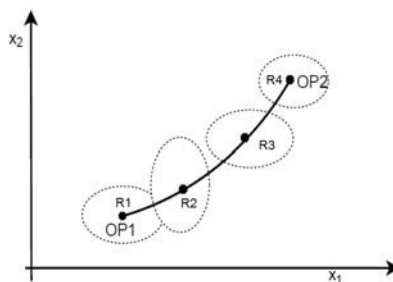
Figure 2.1 Transition of piecewise linear models.

These transitions can be adjusted by higher order TS models, with rules representing each linear region of an overall nonlinear model as an *LTV* (linear time variant) system. In this case, the estimated output is either a mean value between the two models without intersection (R3 and R4) or the weighted model response of two close operational points (R1 and R2).

2.1. Takagi-Sugeno Fuzzy Models

Takagi-Sugeno fuzzy models can be described for a multiple input single output (MISO) system as a set of $i = (1, 2, 3 \dots l)$ rules in the following form:

$$R_i: \text{IF } u_1 \text{ is } A_1 \text{ AND } u_2 \text{ is } A_2 \text{ AND} \dots \text{AND } u_n \text{ is } A_n \text{ THEN } \hat{y}_i(k) = f(u_1, u_2, \dots, u_n) \qquad (3)$$

where $u_1 \dots u_n$ are inputs of the fuzzy model, $A_1 \dots A_n$ are fuzzy sets represented by membership functions which weights a crisp input u_n in a degree of fulfilment such as $w_{jn}(u_n): \mathbb{R} \to [0,1]$; $\hat{y}_i$ is the linear model of the ith rule. The resultant model *(4)* is given by a *t-norm* operation of the inputs and the normalized weights *(5)* in the composition of the *l* rules.

$$\hat{y} = \Phi_1 y_1 + \Phi_2 y_2 + \dots + \Phi_l y_l \qquad (4)$$

$$\Phi_i = \frac{w_i}{\sum\limits_{j=1}^{l} w_j} ; \qquad \sum\limits_{i=1}^{l} \Phi_i = 1 \qquad (5)$$

For a piecewise step response model *(6)* with a horizon *H*, the weighting function Φ is applied by means of superposition principle over the step response coefficients of each linear model.

$$\hat{y}(t+k) = \sum\limits_{m=k+1}^{H-1} S_m^* \Delta u(t+k-m) + S_T^* u(t+k-H) ; \quad S_m^* = \sum\limits_{j=1}^{l} \Phi_j S_j \qquad (6)$$

Here S^* is the weighted toeplitz matrix of the step response coefficients and S_j is the matrix representing the jth sub-model of the PWL or PWA (Piecewise Affine) system.

The PWA representation of a discrete state space model *(7)* is similar to the above step response model. The state matrix A^* and the input matrix B^* are also weighted by the function Φ, while A_j and B_j matrix *(8)* represents the local linearization at the jth equilibrium point of a nonlinear model.

$$\begin{cases} \hat{x}(k+1) = A^* \hat{x}(k) + B^* u(k) \\ \hat{y}(k) = C^* \hat{x}(k) + D^* u(k) \end{cases} \qquad (7)$$

$$A^* = \sum\limits_{j=1}^{l} \Phi_j A_j ; B^* = \sum\limits_{j=1}^{l} \Phi_j B_j ; C^* = \sum\limits_{j=1}^{l} \Phi_j C_j ; D^* = \sum\limits_{j=1}^{l} \Phi_j D_j \qquad (8)$$

In Figure 2.2, a simple representation of the PWL Model Predictive control is placed in the sense of an internal model structure (Roubos et al, 1999). The bias correction ($y(t) - \hat{y}(t)$) is used to prevent modeling errors and in the estimation of unmeasured disturbances. The sub-models are composed by means of fuzzy reasoning to a single

model. The predicted output vector is sent to the optimizer and the quadratic convex function is solved for $\Delta u(t)$.
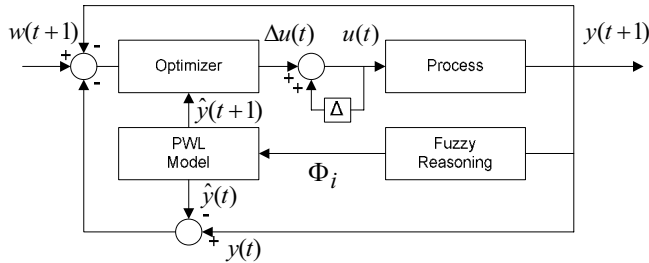
Figure 2.2 PWL Model Predictive control scheme.

3. Control Problem

A non-isothermal Continuous Stirred Tank Reactor (CSTR) was utilized as a control benchmark problem (see Figure 3.1a) for the PWL MPC. The nonlinear system presented by Luyben (1995) consists of an irreversible, exothermic reaction $(A \rightarrow B)$ carried out in a perfectly mixed CSTR. The three state model is given by the reactor temperature, $T(K)$, the reactor feed concentration, $CA(kmol.m^{-3})$, and the jacket reactor temperature $T_j(K)$. The problem consists in controlling the reactor temperature manipulating the makeup jacket flow, $F_j(m^3/hr)$, bringing the process to a desirable product concentration, $CB(kmol.m^{-3})$, value.

The motivation in using this example is due to the characteristics it presents with the output multiplicity behavior of the system. Here, a single input can lead the process to three different output values as shown in Figure 3.1b. For this system, the intermediate region is unstable in open-loop. For the open-loop case, depending on the input signal, the system can go to an ignition temperature (higher temperature region) or to an extinction temperature (lower temperature region) (Bequete, 2007).

(a)   (b)
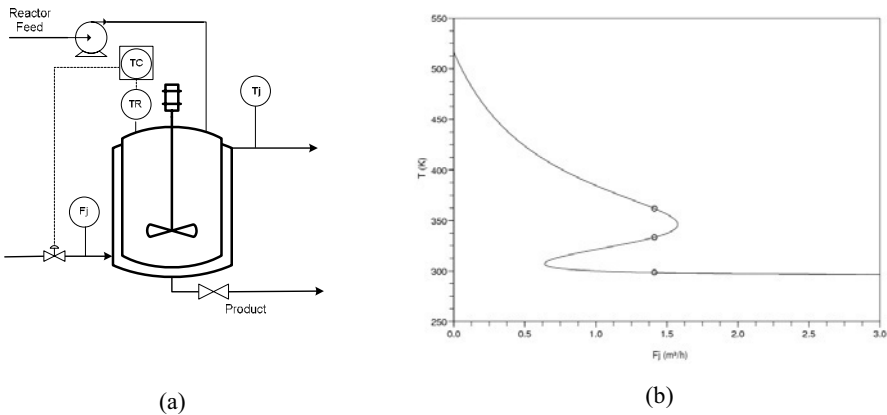
Figure 3.1 (a) Representation of the Continuous Stirred Tank Reactor (CSTR) problem; (b) Multiple steady-state behavior for the proposed control problem.

4. Results

Two PWL MPC controllers were designed for the temperature control of the proposed problem starting with $F_{jss}=1.41m^3/hr$ at a sampling time of $0.08hr$; A dynamic matrix controller (DMC) was developed for the stable low temperature region ($CA_{ss}=7.6kmol/m^3$, $T_{ss}=298.42K$) and a state space model predictive controller (SSMPC) for the intermediate unstable region ($CA_{ss}=3.92kmol/m^3$, $T_{ss}=333.34K$). For the comparisons, a standard MPC of each type and a PI (proportional-integral) controller optimally tuned for the reference trajectory were utilized.

For the standard DMC algorithm, a single step response is taken in the region *R2*. The PWL DMC takes two additional regions into account covering a large area of operation given by *R1* and *R3*. For the SSMPC the model *Q1* is taken by linearization of the nonlinear model in the limits of $T=333.05K$. The PWL SSMPC takes an additional model *Q2* in the limits of the temperature $T=335.83K$. The fuzzy sets for the PWL controllers were chosen heuristically and are given in Figure 4.1.

In both simulations, the system was submitted to setpoint changes making the reactor temperature pass along the regions mentioned above. Control movements were penalized by a weighing factor $R=0.01$ for both configurations of MPC. SSMPC was subject to constraints in the changing of control signal ($\Delta u \leq 0.425m^3/hr$) preventing large temperature overshoots.

(a)　　　　　　　　　　　　　　　(b)
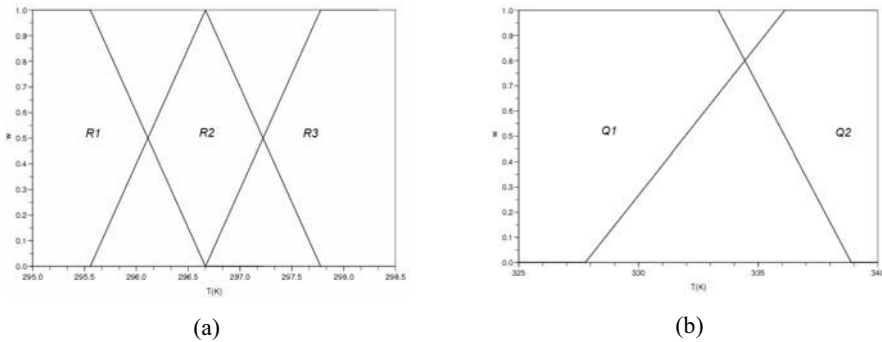
Figure 4.1 (a) Fuzzy sets for the PWL DMC; (b) Fuzzy sets for the PWL SSMPC.

(a)　　　　　　　　　　　　　　　(b)
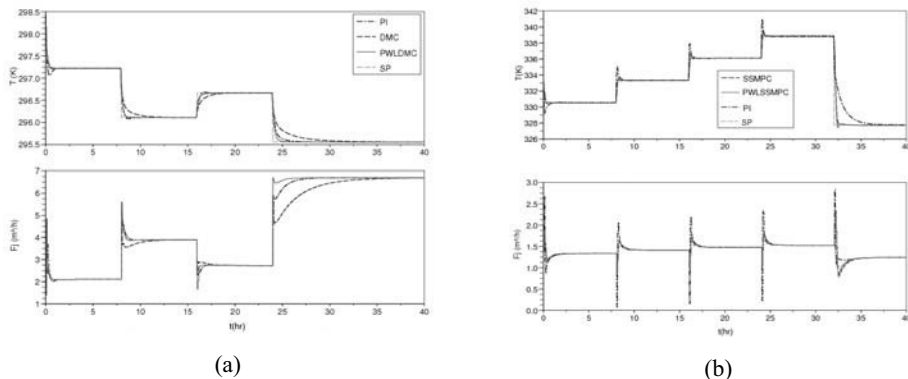
Figure 4.2 Closed-loop behavior of reactor temperature: (a) PWL DMC controller (Hc=2, Hp=4); (b) PWL SSMPC controller (Hc=10, Hp=15).

The closed-loop behavior of the system is given in Figure 4.2. Results showed that, for the lower temperature region (see Figure 4.2a), the PWL DMC had a better response due to the correction provided by the multiple sub-models approach on the final gain, leading the controller to a faster response. In Figure 4.3b the PWL SSMPC controller showed a slightly better response, the similar behavior were due to the high interpolation between the two models *Q1* and *Q2*. A small interpolation of the models was found to be unsatisfactory, due to the high nonlinearity and instability of that region. Furthermore, the great advantage of the proposed approach is the possibility to walk in the entire trajectory of the intermediate region, which cannot be done with a single model, depending on the nature of the operating point. It is shown that the PI controller is able to perform reasonably well, at the great cost of control moves and overshoot.

5. Conclusions

The PWL MPC based on a fuzzy logic description is introduced and illustrated with success in nonlinear process with output multiplicity. The overall performance of the controller depends not only on the number of local models considered, but also how the membership function is defined. Perhaps complex systems will require an optimization procedure for designing the membership region for each model. The structure presented herein is valid and can be easily adapted to any linear MPC strategy. The controller can be easily adapted to several cost function norms and can incorporate robustness constraints with easy.

References

B. Bequette, 2007, Non-Linear Model Predictive Control: A Personal Retrospective, The Canadian Journal of Chamical Engineering, 85, 408-415.

J. Espinosa & J. Vandewalle, 1998, Predictive Control Using Fuzzy Models Applied to a Steam Generating Unit, Proceedings of the Third International FLINS Workshop.

J. Jang, 1993, Anfis: Adaptive network based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics, 23, 665-685.

J. Maciejowski, 2002, Predictive Control with Constraints, Tokyo Denki University Press.

J. Roubos, S. Mollov, R. Babuska & H. Verbruggen, 1999, Fuzzy model-based predictive control using Takagi-Sugeno models, International Journal of Approximate Reasoning, 22, 3-30.

M. Fischer, M. Schmidt & K. Biasizzo, 1997, Nonlinear Predictive Control Based on the Extraction of Step Response Models from Takagi-Sugeno Fuzzy Systems, Proceedings of the American Control Conference, 5, 2878-2882.

P. Marusak, 2007, Advantages of an Easy to Design Fuzzy Predictive Algorithm: Application to a Nonlinear Chemical Reactor, Proceedings of the International Multiconference on Computer Science and Information Technology, 327 – 336.

R. Babuska, J. Roubos & H. B. Verbruggen, 1998, Identification of MIMO systems by input-output TS fuzzy models, IEEE World Congress on Computational Intelligence, 1, 657-662.

T. Takagi & M. Sugeno, 1985, Fuzzy identification of systems and its applications to modeling and control, IEEE transactions on systems, man, and cybernetics, 15, 116-132.

W. Luyben, 1996, Process Modeling, Simulation and control for Chemical Engineers, 2ed., McGraw-Hill, Singapore.

Y. Huang, H. Lou, J. Gong & T. Edgar, 2000, Fuzzy model predictive control, IEEE Transactions on Fuzzy Systems, 6, 665-678.

Optimal Operation of an Industrial Smelter Furnace Off-Gas Cleaning System

Antonio C. B. Araujo[a], Romildo P. Brito[a], Helen Shang[b]

[a]*Federal University of Campina Grande, Av. Aprigio Veloso 882, Campina Grande PB 58429-900, Brazil*
[b]*Laurentian University, 935 Ramsey Lake Road, Sudbury ON, P3E 2C6, Canada*

Abstract

Off-gas cleaning systems extract and treat hazardous emissions, and ensure that smelter operations are in accordance with environmental and industrial hygiene regulations. To this end, it is paramount that a well designed control structure be incorporated into the system. We first approach the problem by conducting a steady-state analysis based on a nonlinear model of the process, where the objective is to achieve safe, clean, and economic operation. Results reveal that this is accomplished by keeping the temperature in the furnace and in the two louvers at their upper bounds (active constraints). The unconstrained variable is found by applying the self-optimizing control technique, and the results indicate that a small loss is acceptable when one of the manipulated variables is fixed. The bottleneck of the system is identified as the fans' discharge pressures when we allow the feed rate as a degree of freedom. A control structure is then designed to keep pressures in the system well within the negative region such that acceptable dynamic performance in face of known, deterministic disturbance is achieved. Nonlinear dynamic simulations showed that very good dynamic performance is achieved with decentralized PI control schemes for both regulatory and supervisory designs.

Keywords: smelter, furnace, off-gas, optimization, control structure.

1. Introduction

Smelter furnaces provide the key operations in obtaining metal products from mineral concentrates. However, a byproduct of the process is the hazardous emission of off-gases which constitute a major source of atmospheric and plant environment pollutions. This gas is ladened with significant concentrations of CO, CO_2, and small amounts of SO_2. Continued efforts on process control and optimization is necessary to further improve the operations of smelter off-gas cleaning systems. In this work, the decision of which variables to control in an industrial furnace smelter off-gas system is addressed, and the issue is to ensure the operational objectives are achieved at the lowest possible cost, while at the same time avoiding atmospheric emissions of hazardous gases. A throughput analysis is carried out to determine where in the process the bottleneck is located, and hence where production rate should set. These decisions support the design of control configurations for the process, both at the regulatory and supervisory layers.

2. Process description

A schematic representation of the process is given in Figure 1. Hot calcines from other parts of the process, together with dusts from the off-gas cyclones and Cottrell plant are

fed to the furnace, to which coke and air are also added. Each furnace is equipped with electrodes to provide the heat required for complex reactions to take place (Celmer et al., 1987). The off-gas generated in the furnace freeboard consisting of CO, CO_2, SO_2, N_2 and O_2, first enters a cyclone to remove fine particles, which are recycled back to the furnace. It then passes a louvre for air cooling and through a fan that discharges the cooled gas into the Cottrell plant. The fans in the gas system are essential in maintaining a slightly negative pressure and suitable gas temperature in the furnace freeboards. Converters post-process the electric furnace product to remove the remainder of the iron as a slag byproduct.
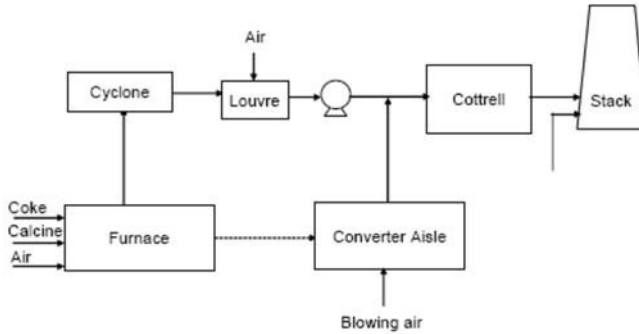
Figure 1: Schematic diagram of an industrial smelter furnace off-gas system.

3. Control structure design of the off-gas cleaning system

In this section, we design a control structure for the process based on the procedure of Skogestad (2004a), starting with a steady-state analysis.

3.1. Definition of optimal operation and degree of freedom analysis

The objective is to use as little power as possible in the two fans, subject to avoiding gas out-leakage. This is equivalent to minimizing the air in-leakage in the furnace while still preventing off-gas from finding a way out of the system. Therefore, the scaled cost function is given by (1).

$$J = pK/(\varepsilon\rho C_q) \ [(P_{fan1} - P_{lv1})W_{fan1} + (P_{fan2} - P_{lv2})W_{fan2}], \qquad (1)$$

where p is the price of electricity; K is a coefficient that accounts for the compressibility of the gas, assumed constant for the range of pressures considered in this work; ε is the fan efficiency, assumed the same for all fans; ρ is the gas density, assumed constant for the range of pressures considered in this work; C_q is a constant that adjusts the dimension of the units used; the other variables are defined below. The constraints to operation are listed in Table 1. The process has 5 manipulated variables - coke added to the furnace (W_{coke}), louver #1 and #2 vane openings (O_1 and O_2), fan #1 and #2 rotation speeds (N_{fan1} and N_{fan2}) - and therefore 5 dynamic degrees of freedom. We also consider 16 candidate measurements: furnace freeboard pressure (P_f) and temperature (T_f); cyclone #1 and #2 pressures (P_{c1} and P_{c2}); louver #1 and #2 pressures (P_{lv1} and P_{lv2}); louver #1 and #2 temperatures (T_{lv1} and T_{lv2}); Cottrell inlet pressure (P_{iCt}) and temperature (T_{iCt}); Fan #1 and #2 outlet pressures (P_{fan1} and P_{fan2}); louver #1 and #2 vane openings (O_1 and O_2); and fan #1 and #2 rotation speeds (N_{fan1} and N_{fan2}). The disturbances in Table 2 include the effects of feed flow rate on the process and changes in the system's structure as represented by the effective in-leakage area in the furnace.

Table 1: Steady-state constraints to the process.

	Constraint description	Bound
P_f	Furnace freeboard pressure [Pa]	< -25
P_{c1} and P_{c2}	Cyclone #1 and #2 pressures [Pa]	< 0
P_{lv1} and P_{lv2}	Louver #1 and # 2 pressures [Pa]	< 0
P_{fan1} and P_{fan2}	Fan #1 and # 2 outlet pressures [Pa]	< 0
P_{iCt}	Cottrell inlet pressure [Pa]	< 0
T_f	Furnace freeboard temperature [K]	< 923.15
T_{lv1} and T_{lv2}	Louver #1 and #2 temperatures [K]	< 643.15
T_{iCt}	Cottrell inlet temperature [K]	< 643.15
O_1 and O_2	Louver #1 and #2 vane openings [%]	[0 100]
N_{fan1} and N_{fan1}	Fan #1 and #2 rotation speeds [rpm]	< 1500

3.2. Primary controlled variable selection

The starting point for the selection of primary (economic) variables is the optimization of the process for the disturbances in Table 2. The model developed by Shang et al. (2008) is implemented and optimizations with fixed W_{coke} are performed. Figure 2 shows the effect of the disturbances on the cost. 3 constraints are always active, namely, T_f (upper bound), T_{lv1} (upper bound), and T_{lv2} (upper bound). As these 3 active constraints must be implemented to ensure optimal operation (Maarleveld and Rijnsdorp, 1970), we are left with 1 degree of freedom.

Table 2: Disturbances to the process.

Disturbance description	Nominal	Disturbance (Δ)
Coke added to the furnace [kg/s]	2	+0.4 (D1) and -0.2 (D2)
Equivalent CO temperature in the furnace [K]	1573.15	+200 (D3) and -200 (D4)
Flow rate from converters [kg/s]	300	+60 (D5) and -60 (D6)
Converter outlet temperature [K]	473.15	+100 (D7) and -100 (D8)
Effective in-leakage area in the furnace [m2]	1.50	+0.75 (D9)
Room temperature [K]	273	-30 (D10) and +30 (D11)

It can be readily pointed out from Figure 2 that the disturbances which have the largest influence on the cost are those that affect the mass balance, i.e., D1 and D2. By increasing the flow rate from the converters the cost is reduced, and vice-versa. This is due to the reduction in the head from the louver to the fan. When the equivalent CO temperature in the furnace increases (D3 and D4), pressure in the furnace freeboard is reduced allowing more air into the system to cool down the off-gas, and vice-versa. The converter temperature (D7 and D8) has no economical effect on the system as it has no influence in upstream units. Disturbance D9 shows that it is "cheaper" to operate the furnace at higher pressures but within a safety margin. The seasonal change in temperature (D10 and D11) affects the economics of the system by allowing more or less cooling load.
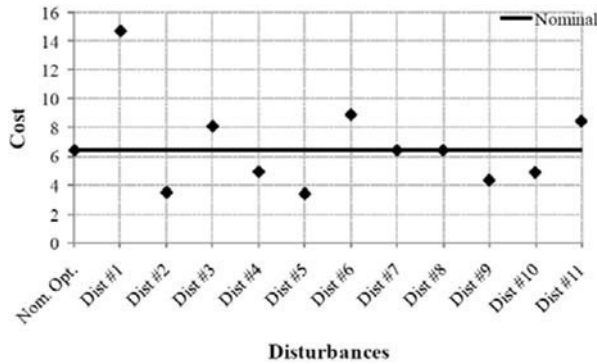
Disturbances

Figure 2: Effect of disturbances on optimal operation.

We now need to find a suitable unconstrained controlled variable, and to this end we use the self-optimizing control technology (Skogestad, 2000) which consists of finding a variable c that when kept at its nominal optimal setpoint minimizes the loss $L = J(c,d) - J_{opt}(d)$, where $J(c,d)$ is the actual value of the cost function for a given c and $J_{opt}(d)$ is the truly optimal value if the process was re-optimized for each disturbance d. The results for the various runs showed that a good candidate would be either N_{fan1} or N_{fan2}.

3.3. Production rate

The maximum throughput analysis determines the evolution of the static behavior of the system as W_{coke} is increased. It is usually optimal to increase the production rate as much as possible because the prices may be such that the actual profit of the enterprise would increase with W_{coke}. To find $W_{coke.max}$ (Mode II), and hence the bottleneck of the process, we use W_{coke} as a degree of freedom and re-optimize the process, using the profit function J' as given in (2), where p_f is the price of the feed.

$$J' = pK/(\varepsilon\rho C_q) \ [(P_{fan1} - P_{lv1})W_{fan1} + (P_{fan2} - P_{lv2})W_{fan2}] - p_f W_{coke} \tag{2}$$

The cost as a function of W_{coke} is shown graphically in Figure 3. The active constraints were found to be the same as per Mode I, i.e., T_f, T_{lv1}, and T_{lv2} are active at their upper bounds. Operation becomes infeasible when $W_{coke} = 3.15$ kg/s since both P_{fan1} and P_{fan2} become active at their respective upper bounds; consequently, production cannot be further increased, and the bottleneck of the process is either P_{fan1} or P_{fan2}. However, when disturbances are considered $W_{coke.max}$ will change accordingly. Assuming the plant is subject to disturbances D3 to D11 in Table 3, the maximum throughput results show that the converters largely affect the steady-state economic operation of the off-gas system, and for these disturbances, the throughput varies largely from $W_{coke} = 2.52$ to 3.78 kg/s.
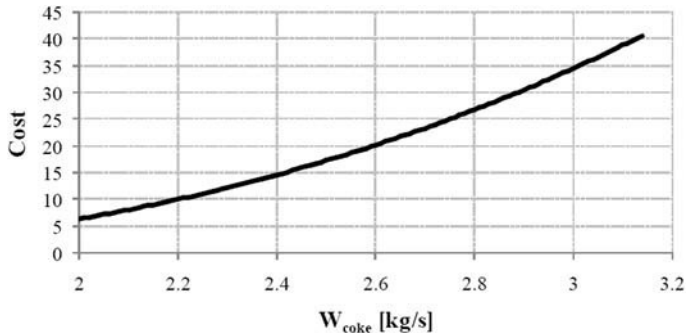
Figure 3: Optimization of the smelter furnace system with variable feed rate.

4. Bottom-up design of the cleaning off-gas system

The main objective of the regulatory layer is to provide sufficient control quality to enable a trained operator to keep the plant running safely without the need for using higher layers in the control system. In addition, the control task at this layer is to prevent the plant from drifting away from its desired operating point on the short time scale. As the smelter furnace off-gas system has no unstable mathematical modes, no stabilization of this kind is indeed necessary. However, as pressure dynamics are generally very fast, drift in these variables due to disturbances is avoided by controlling pressure at selected locations in the plant. In Mode I, P_f is selected to be controlled using either N_{fan1} or N_{fan2}, this minimizes the impact of disturbances in the primary controlled temperatures because of the direct relation between these two variables. As for Mode II, P_f is also controlled, and we use W_{coke} as manipulated variable; in addition, we control P_{fan1} and P_{fan2} using N_{fan1} and N_{fan2}, respectively. The intended aim of the supervisory control layer is to keep the active constraints and unconstrained controlled variables at their constant optimum setpoints. For the unconstrained variable in Mode I, N_{fan1} is fixed at its nominal optimum set point. For Mode II, T_f is controlled using as manipulated variable the setpoint of the furnace pressure controller, $P_{f,sp}$. These two strategies favor optimal economic performance with minimal losses. Moreover, for both modes of operation T_{lv1} and T_{lv2} are controlled using O_1 and O_2, respectively. The production rate manipulator is selected to be W_{coke}, where it is fixed in Mode I and adjusted to give the desired maximum throughput in Mode II. The resulting PI controllers are tuned using the SIMC tuning rules described in Skogestad (2004b). Note that optimal operation in both modes always lies at process constraints. However, during transients these constraints may eventually be violated for very short periods of time when using conventional control configurations, since PID controllers do not account for constraints in their algorithms. Nonetheless, the variables associated to these constraints must always return to their bounds irrespective of disturbances, and this can always be achieved by adding integral action to the controllers. One way to minimize these dynamic violations is to resort to multivariable controllers which explicitly make up for constraints in their formulations like, for instance, MPC controllers.

5. Validation of the proposed control structure

We here perform dynamic simulations to validate the proposed control structures for the smelter furnace off-gas system when facing fast changing disturbances. For the sake of brevity, only simulations involving Mode I with disturbances D1 and D2 are considered. Figure 4 illustrates the responses of the system where it can be seen that the proposed control structure is effective in rejecting disturbances, showing quick dynamic response, with pressures strictly within bounds.
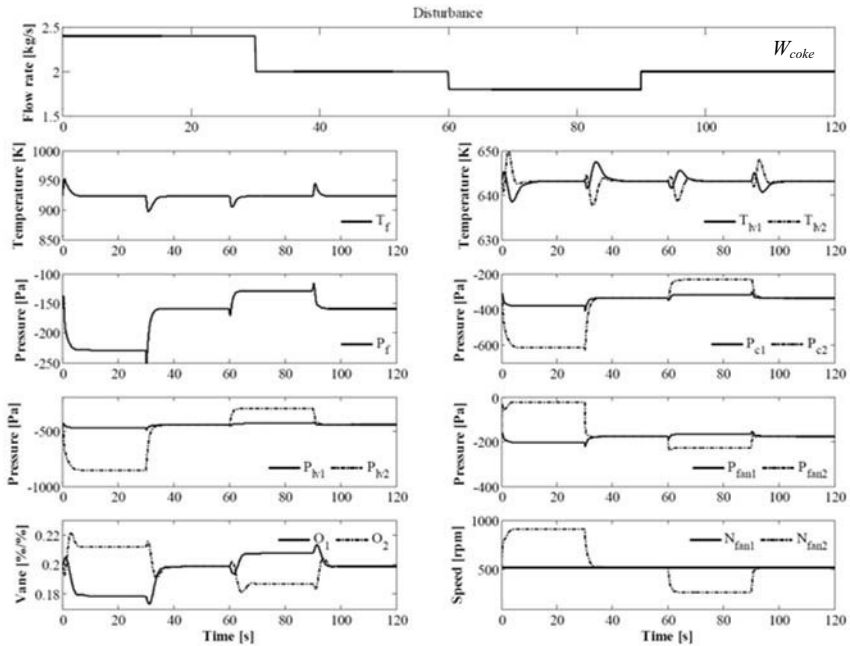
Figure 4: Responses for disturbances in W_{coke} (top of the figure) in Mode I.

The responses for Mode II are not shown but they yield good dynamic performance, with pressures well within their negative ranges.

6. Conclusion

The use of a systematic procedure to the design of a simple decentralized control structure shows it is possible to operate an industrial smelter furnace off-gas cleaning system in a (near) optimal economic fashion that entirely complies with environmental regulations by avoiding out-leakage of hazardous off-gases to the atmosphere. The resulting dynamic performances for both modes of operation ratify the topnotch efficiency of the proposed control configurations when facing large, fast disturbances affecting the process.

References

Celmer, R. S., Kaiura, G. H., Toguri J. M., 1987, Chemical reactions during the electric smelting of nickel-copper calcines, Canadian metallurgical Quarterly, 26(4), 277-284.

Maarleveld, A., Rijnsdorp, J. E., 1970, Constraint control on distillation columns. Automatica, 6, 51-58.

Shang, H., Megan, D., Nelson, P., Salt, B, 2008, Dynamic modelling of an industrial smelter furnace and converter off-gas system, American Journal of Environmental Sciences, 40(1), 22-30.

Skogestad, S., 2000, Plantwide Control: The search for the self-optimizing control structure, Journal of Process Control, 10, 487-507.

Skogestad, S., 2004a, Control structure design for complete chemical plants, Computers and Chemical Engineering, 28, 219-234.

Skogestad, S., 2004b, Simple analytic rules for model reduction and PID controller tuning, Modeling, Identification and Control, 25(2), 85-120.

Optimal Operation of an Industrial PVC Dryer

Antonio C. B. Araujo[a], Jose J. N. Neto[a], Helen Shang[b]

[a]*Federal University of Campina Grande, Av. Aprigio Veloso 882, Campina Grande PB 58429-900, Brazil*
[b]*Laurentian University, 935 Ramsey Lake Road, Sudbury ON, P3E 2C6, Canada*

Abstract

In this work, we devise a control structure architecture for an industrial polyvinyl chloride (PVC) dryer currently operating at Braskem Company, Brazil. The motivation is the optimally economic operation of the process as well as the prompt rejection of disturbances at lower layers in the control hierarchy. Optimization of a simplified model of the process for various important disturbances indicates that it is optimal to control the temperature level of the utilities serving the dryer and the final PVC moisture contents at their respective upper bounds. In addition, acceptable economic loss is achieved by keeping the flow of air to the dryer at a fixed optimum setpoint. Almost perfect indirect control of the final PVC moisture is achieved by tightly controlling a temperature difference in the dryer. The resulting decentralized control configuration leads to good dynamic performance for important disturbances affecting the system.

Keywords: drying, PVC, optimization, control structure.

1. Introduction

Fluidized bed dryers are widely used in both continuous and batch processes by the food, pharmaceutical, agricultural, and chemical industries. However, systematic design of control structures is still lacking. This leads to the fact that literature on control of dryers is surprisingly scarce, in addition to being based on heuristics arguments. We here name only a few results found in most relevant papers. Mujumdar (2006) described control strategies for fluidized bed dryers, and quotes that the most adequate is when the exhaust-air temperature is controlled by the inlet air temperature. Harbert (1974) and Alden *et al.* (1988) studied a technique where the temperature difference between the material being dried and the wet-bulb temperature is the selected controlled variable since it infers the end-point of drying. Abdel-Jabbar *et al.* (2005), on the other hand, proposed to control the final moisture content indirectly estimated via a linearly designed Kalman filter. Panda and Rao (1994) derived and compared the performance of internal model (IMC) and dynamic matrix (DMC) controllers in an experimental apparatus for fluidized-bed drying of sand. Most of the papers refer to vertical pneumatic fluidized bed dryers in which the sole heating media is the fluidizing air. In this paper, a different arrangement is considered for the drying of PVC where heating elements are inserted in the bed aiming at reducing capital costs.

2. Description of the industrial unit

Figure 1 depicts a schematic of the PVC drying operation. After polymerization, a mixture containing 30% (wt/wt) of PVC and 70% (wt/wt) water is sent to centrifuges to mechanically reduce the moisture of the slurry to 20-30% wt/wt water content. The resin is then fed into a fluidized bed dryer to remove the residual water. The energy

required by the drying process is supplied by hot air blown to the bottom of the dryer as well as by hot water circulating inside the bed through a series of heat exchangers.

3. Modeling of an industrial PVC dryer

We base the model developed in this paper on the work by Levi and Borde (1999) and Groenewold and Tsotsas (2007) due to its simplicity and accuracy against plant data. The process is described by a two-phase model. Considering a single cell j, we assume that the gas phase is forced vertically to flow through the fluidized particles in a plug-flow pattern. We then divide the gas path into M sections along the cell j where each section i is modeled as a CSTR. In other words, the model is distributed for the gas phase and lumped for the particle phase in a given cell.
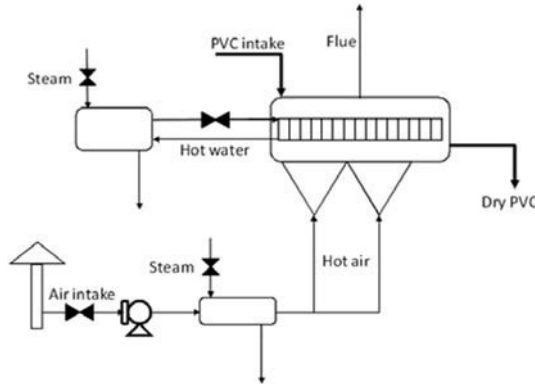
Figure 1: Schematic of the continuous PVC fluidized bed drying process.

A comparison between the proposed model and plant data was performed and the results showed very good agreement.

4. Control structure design of the PVC dryer

In this section, we design a control structure for the process based on the procedure of Skogestad (2004a), starting with a steady-state analysis.

4.1. Definition of optimal operation and degree of freedom analysis

The objective is to reduce the cost of energy as much as possible, which is equivalent to minimizing the amount of steam used to heat up the air and hot water, while still delivering the final PVC at the right moisture content specification. Therefore, the cost function J to be minimized is given by (1).

$$J = \frac{8000}{D_{scl}} \left(F_G \int_{T_{room}}^{T_{G,in}} C_{PG} dT + m_W \rho_W \int_{T_{W,out}}^{T_{W,in}} C_{PL} dT \right) \tag{1}$$

where F_G is the air flow rate; T_{room} is the room temperature; $T_{G,in}$ is the inlet air temperature; $T_{W,in}$ is the inlet hot water temperature; $T_{W,out}$ is the outlet hot water temperature; m_W is the hot water flow rate; ρ_W is the liquid water density; $C_{PG} = C_{PW} + Y_G C_{PA}$ is the gas phase specific heat (C_{PW} and C_{PA} are the specific heat of water vapor and the specific heat of air, respectively for which equations were found experimentally from plant experiments); and C_{PL} is the liquid water specific heat. We here consider operation over a period of 8000 h/year. For optimization purposes the cost as given by (1) is properly scaled as given by the scaling factor D_{scl}. The process is subject to the constraints in Table 1. These constraints reflect two unmodeled features of the process:

The lower bound on the final particle moisture content, $X_{S,out}$, prevents electrostatic charge formation, and the bounds on the air flux ($f_{G,j}$) in each cell j refer to maximum and minimum fluidization velocities through the maximum ($f^{max}_{G,j}$) and minimum ($f^{min}_{G,j}$) air fluxes, for which equations are found in Howard (1993). The process has 5 manipulated variables - PVC feed rate in dry basis (F_S), inlet air flow rate in dry basis (F_G), hot water flow rate (m_W), inlet air temperature ($T_{G,in}$), and inlet hot water temperature ($T_{W,in}$) - and therefore 5 dynamic degrees of freedom. We also consider 10 candidate measurements: gas temperature at the bottom of cell #1 ($T_{G1,1}$), #2 ($T_{G1,3}$), and #3 ($T_{G1,4}$) and respective differences $T_{G1,1}$ - $T_{G1,3}$; $T_{G1,1}$ - $T_{G1,4}$; and $T_{G1,3}$ - $T_{G1,4}$; outlet gas temperature ($T_{G,out}$); outlet hot water temperature ($T_{W,out}$); inlet air flow rate (F_G); and hot water flow rate (m_W). Disturbances in Table 2 represent changes in the inlet PVC.

Table 1: Constraints to the process.

Constraint description	Bound
Final particle moisture content in dry basis [kg/kg]	$0.0018 \leq X_{S,out} \leq 0.0020$
Air flux in cell j [kg/(m^2 s)]	$f^{min}_{G,j} \leq f_{G,j} \leq f^{max}_{G,j}$
Inlet air temperature [K]	$303.15 \leq T_{G,in} \leq 383.15$
Inlet hot water temperature [K]	$303.15 \leq T_{W,in} \leq 383.15$
Inlet air flow rate in dry basis [kg/h]	$22370 \leq F_G \leq 78300$
Hot water flow rate [m^3/h]	$0 \leq m_W \leq 300$

4.2. Primary controlled variable selection

The starting point for the selection is the optimization of the process for the disturbances in Table 2. The proposed model is implemented and optimizations with fixed F_S are performed. The results show that 3 constraints are always active at their upper bounds, namely, $X_{S,out}$, $T_{G,in}$, and $T_{W,in}$. $X_{S,out}$ is active since it is always optimal to operate the plant at the minimum specification. $T_{G,in}$ and $T_{W,in}$ are active because drying processes are basically favored by high energy transfer rates. Therefore, what is left for the optimization is to minimize the flow rate of utilities to trade off cost reduction with feasibility, while satisfying the specification on $X_{S,out}$. As these 3 active constraints must be implemented to ensure optimal operation (Maarleveld and Rijnsdorp, 1970), we are left with 1 degree of freedom.

Table 2: Disturbances to the process.

Disturbance description	Nominal	Disturbance (Δ)
Particulate phase mass flow rate in dry basis [kg/h]	7370	+1630 (D1) and -2370 (D2)
Initial particle moisture content in dry basis [kg/kg]	0.233	+0.282 (D3) and -0.043 (D4)
Initial particulate phase temperature [K]	335	+23 (D5) and -2 (D6)

It can be seen from Figure 2 that the disturbances which most influence the cost is when $X_{S,in}$ is increased because it is rather difficult to remove the moisture inside the particle since more air would be necessary, and this increase in air flow rate will eventually hit the upper bound on f_G. Therefore, hot water is used to provide the complementary energy requirements, which is a more expensive heat source.

We now need to find a suitable unconstrained controlled variable, and to this end we use the self-optimizing control technology (Skogestad, 2000) which consists of finding a variable c that when kept at its nominal optimal setpoint minimizes the loss $L = J(c,d) - J_{opt}(d)$, where $J(c,d)$ is the actual value of the cost function for a given c and $J_{opt}(d)$ is the truly optimal value if the process was re-optimized for each disturbance d.
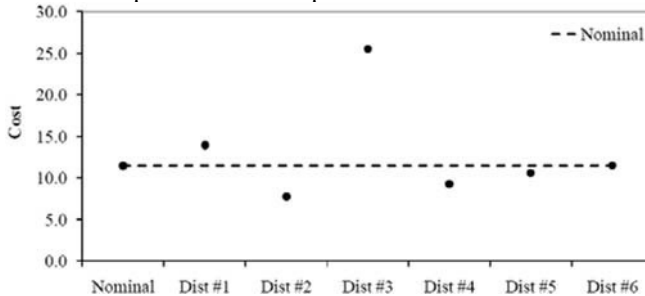
Figure 2: Effect of disturbances on optimal operation.

The results for the various runs show that operation is infeasible for any candidate controlled variable. We then "back off" the variables' optimal setpoints to resolve this problem until finding a suitable value for the selected candidate variables such that operation is always feasible, at the expense of a loss penalty. Two very attractive candidate variables are F_G and m_W since it is always a good practice to operate with minimal manipulation handling. The only value for F_G such that operation is still "cheap" would be the one calculated for disturbance D3, i.e., $F_G = 26300$ kg/h, because anything smaller would result in infeasible operation since m_W is at its maximum. m_W can only assume values such that $m_W \leq 5.32$ m³/h, otherwise operation would become infeasible for D2 as F_G is already at its lower bound. The seemingly optimal value would then be $m_W = 5.32$ m3/h; however, under these conditions it is certain that operation for D3 will be infeasible since with such a small flow of hot water, either the upper bound on f_G or F_G would be violated. Therefore, F_G is chosen as the unconstrained variable.

Since $X_{S,out}$ is difficult to measure, it is necessary to find another suitable active constraint. The idea is to determine the value of a selected variable such that, when kept constant, minimizes the deviation in $X_{S,out}$ from its optimal value $X^{opt}_{S,out}$ when different disturbances are considered. The results of the optimizations show that only the candidate $T_{G1,3}$ - $T_{G1,4}$ gives feasible operation for the entire set of disturbances. Hence, it is select as the new active constraint.

4.3. Production rate

The maximum throughput analysis determines the evolution of the static behavior of the system as F_S is increased. It is usually optimal to increase the production rate as much as possible because the prices may be such that the actual profit increases with F_S. To find $F_{S,max}$ (Mode II), and hence the bottleneck of the process, we use F_S as a degree of freedom and re-optimize the process, using the profit function J as given in (1) (see Figure 3). Note that the active constraints are the same as per Mode I. When $F_S = 12370$ kg/h, f_G reaches its upper bound, and above $F_S = 20810$ kg/h operation is no longer feasible because m_W reaches its upper limit; therefore, m_W is the bottleneck of the process. However, when disturbances are considered $F_{S,max}$ reduces accordingly. By considering the worst case scenario, in which $X_{S,in}$ is increased by 10%, the ultimate throughput is found at $F_{S,max} = 18400$ kg/h.
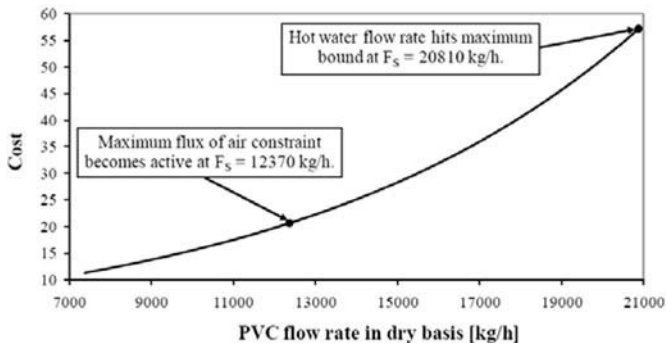
Figure 3: Optimization of the smelter furnace system with variable feed rate.

5. Regulatory and supervisory control structures design

The main objective of the regulatory layer is to provide sufficient control quality to enable a trained operator to keep the plant running safely and to prevent the plant from drifting away from its desired operating point. As the PVC dryer has no unstable mathematical modes, no stabilization of this kind is necessary. To reduce drift, the air blower vane downstream the unit controls pressure, and $T_{G,in}$ and $T_{W,in}$ are kept constant. The intended aim of the supervisory control layer is to keep the active constraints and unconstrained variables at their optimum setpoints. For this, $T_{G1,3}$ - $T_{G1,4}$ is used to indirectly control $X_{S,out}$. The resulting PI controllers are then properly tuned using the SIMC tuning rules described in Skogestad (2004b).

6. Validation of the proposed control structure

For Mode I of operation, we consider an increase of 0.1 kg/kg in $X_{S,in}$. As for Mode II, the disturbance consists of a 10 % step increase in $X_{S,in}$. Figures 4 and 5 illustrates the responses of the system, where it can be seen that disturbances can be quite effectively rejected with few oscillations and overshoots, as well as small offset in $X_{S,out}$.

7. Conclusion

The general guidelines to optimally operate a PVC dryer of the kind discussed in this work can be summarized as: operate the dryer at the maximum temperature level, i.e., with the temperature of the inlet air and hot water at their upper limits, and adjust the air flow rate to minimize the flow rate of hot water, while keeping the final PVC moisture content as close as possible to the upper specification. Having said that, the resulting control structure successfully fulfills the task of delivering PVC at the right moisture content while operating the unit at the lowest possible cost.

References

Abdel-Jabbar, M. N., Ali Jumah, R. Y., and Al-Haj Ali, M. Q., 2005, State estimation and state feedback control for continuous fluidized bed dryers, Journal of Food Engineering, 70, 197–203.

Alden, M., Torkington, P. and Strutt, A. C. R., 1988, Control and instrumentation of a fluidized-bed drier using the temperature-difference technique. I. Development of a working model, Powder Technology, 54, 15–25.

Groenewold, H. and E. Tsotsas, 2007, Drying in fluidized beds with immersed heating elements, Chemical Engineering Science, 62, 481–502.
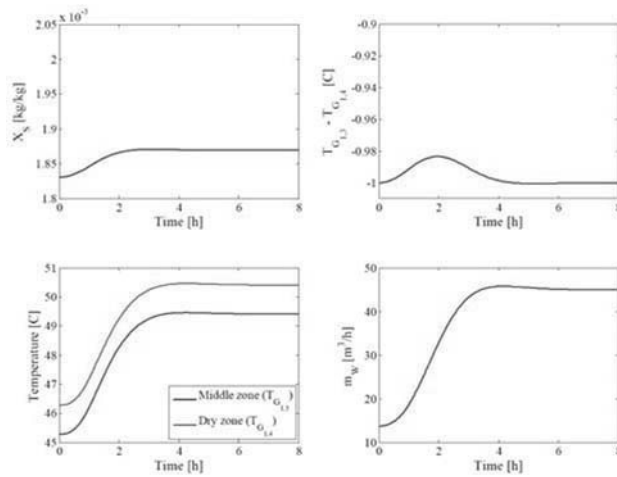
Figure 4: Responses for a disturbance in $X_{S,in}$ in Mode I.
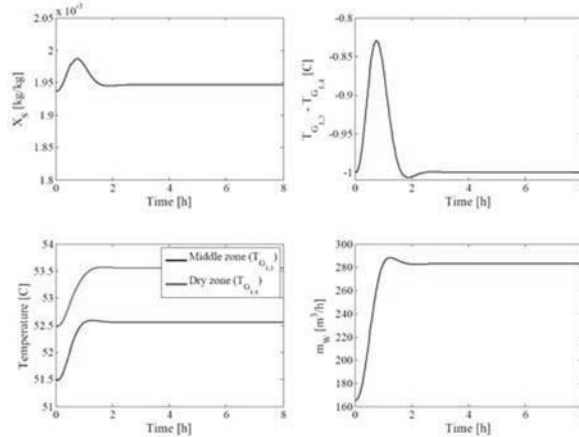
Figure 5: Responses for a disturbance in $X_{S,in}$ in Mode II.

Harbert, F. C., 1974, Automatic control of drying processes moisture measurement and control by the temperature difference method. Chemical Engineering Science, 29,888–890.

Howard, J. R., 1993, Fluidized Bed Technology: Principles and Applications, Adam Hilger.

Levi, A. and I. Borde, 1999, Steady state one dimensional flow model for a pneumatic dryer, Chemical Engineering Processing, 38(2), 121–130.

Maarleveld, A., Rijnsdorp, J. E., 1970, Constraint control on distillation columns. Automatica, 6, 51-58.

Mujumdar, A. S., 2006, Handbook of Industrial Drying, 3rd ed., CRC.

Panda, R. C. and Rao, V. S. R., 1994. Model based control of a continuous fluidized bed dryer. In Proceedings of the third IEEE conference on control applications, 507–511, Glasgow, Scotland.

Skogestad, S., 2000, Plantwide Control: The search for the self-optimizing control structure, Journal of Process Control, 10, 487-507.

Skogestad, S., 2004a, Control structure design for complete chemical plants, Computers and Chemical Engineering, 28, 219-234.

Skogestad, S., 2004b, Simple analytic rules for model reduction and PID controller tuning, Modeling, Identification and Control, 25(2), 85-120.

Optimal Scheduling of a Multiproduct Continuous Paper Plant

Pedro P. Mariano,[a,b] Pedro M. Castro,[b] Ana P.F.D. Barbosa-Póvoa[a]

[a]*CEG-IST, Universidade Técnica de Lisboa, 1049-001 Lisboa, Portugal*
[b]*DMS, Inst. Nacional Engenharia Tecnologia e Inovação, 1649-038 Lisboa, Portugal*

Abstract

This paper addresses the solution of a real-life problem from a multiproduct paper plant, involving the scheduling of three machines in parallel for a time horizon of roughly one month. A decomposition strategy is proposed to reduce the complexity where orders are aggregated according to the product, density, size and due date. Furthermore, they are allocated to a particular machine, enabling the solution of three separate sequencing problems instead of one very large scheduling problem. Three alternative formulations from the literature are evaluated, with the results showing that it is better to use a unit-specific continuous-time formulation with 3-index binary variables. This is then used to find a better solution to the problem by releasing the order-unit assignments for a subset of the orders.

Keywords: Continuous-time, decomposition strategy, parallel machines, changeovers.

1. Introduction

Scheduling is concerned with the allocation of resources over time so has to execute the processing tasks required to manufacture a given set of products. Due to its critical role, this type of activity plays an important part in any organization. In particular, there is the case of industries of the production/transformation sectors with considerable amounts of inventory where an optimization of the associated schedule is required. Pulp and paper is a perfect example, with the Setubal paper plant of Portucel Soporcel, located in Portugal, providing the subject of this paper.

The plant's Production Department plans the budget for production and transformation, which includes cut, cut size and rewind, for a time horizon of one year. This plan is strongly influenced by the market behavior and the data acquired from the clients. Considering this plan and the orders received, the Sales & Marketing Department conceives the schedule for each month. A generic 30-day schedule (each with 3 daily shifts) features 22 changeovers between different products, distributed over the 3 paper machines in the following way: 12 in MP1, 5 in MP2, and 5 in MP3.

This paper aims to optimize the schedule of the plant by minimizing the total production time and the number of changeovers.

2. State-of-the-art

Scheduling has received considerable attention by the scientific community in the last twenty years, in large part due to the vital role it plays in the operations management area. Concerning time representation, two main types of formulations can be identified. Discrete time formulations (DT), where the time horizon is divided into a number of

intervals of equal duration, and continuous time formulations (CT) where the duration of the time intervals is going to be determined by the optimization solver.

The DT models, firstly based on a State-Task (STN) and later on a Resource-Task Network representation of process and/or production recipe, appeared earlier than their continuous-time counterparts. However, the latter have received most of the attention in the last decade (Méndez et al., 2006). Nowadays, researchers are mostly focusing on developing unit-specific approaches (Castro et al. 2006, Liu and Karimi, 2007, Shaik and Floudas, 2008, Castro and Novais, 2008) that can be as general as those based on global event points, since the former can be significantly more efficient computationally. In particular, multistage plants are an example of success. Overall, it is clear that there is not a single best approach for all problem types and researchers are still drawing the map that will tell which model to use as a function of the characteristics of the real problem. This paper evaluates the performance of two models (CT3I and CT4I) proposed by Castro et al. (2006) on a real-life industrial case.

3. Problem definition

In this paper, the short-term scheduling problem of a single stage, multiproduct continuous plant is considered. This is formed by a set of three machines ($m \in M$) that have to process within a month, 420 orders ($i \in I$) of 12 different products (P1-P12). Given are also the duration of the processing tasks and cleaning tasks, the due dates and the amounts to process. The objective is to minimize the total production time in all machines, which maximizes productivity by reducing the number of changeovers, each lasting between 10 to 30 minutes. It is required to meet (hard constraints) the due dates for all products as well as the maximum availability for storage in the warehouse.

4. Solution strategy

The original number of orders (420) is well above the current capabilities of state-of-the-art scheduling models, so a two-stage solution method (figure 1) was applied to reduce the complexity.

In the first stage, orders belonging to the same product and sharing the density (g/m^2), size and due date, were placed in the same group. This reduced the number of orders to 191, still too high to start to solve the current problem. Thus another aggregation step was needed, which consisted on grouping orders with total duration as close as possible to a day of production. The outcome was the reduction of the number of orders to 73, an acceptable value.

In the second stage, we assume a fixed product-unit assignment, which is stated by the Production and Sales & Marketing departments. In this way, orders belonging to P1-P6 are processed in MP1; orders belonging to P7-P9 in MP2; and those belonging to P10-P12 in MP3.

Overall, a trade-off is established between flexibility (more orders to process) and computer effort (less orders). Note that order aggregation and fixed order-units assignments will almost certainly remove the optimal solution from the feasible space. More importantly, the scheduling problem has been replaced by sequencing problems, which will be solved either simultaneously or separately.

5. Continuous-time model selection

The number in CT3/4I indicates the indices in the binary variables. Formulation CT4I identifies the execution of order i followed by the required changeover for order i' to immediately follow in unit m at event point t. This approach allows to model

changeovers explicitly. CT3I, on the other hand, identifies the execution of order i in unit m at time t, and handles changeovers implicitly through the model constraints. This is the main conceptual difference between the two formulations, which is reflected in the size of the mathematical problem by roughly one order of magnitude in the number of variables.

To evaluate which formulation is the most efficient for the present problem, a simplified version of the industrial problem was solved. It featured a half-month time horizon and 36 orders, corresponding to about 8500 tons of paper. One important decision concerns the number of event points to consider on each time grid, one per machine. Usually, an iterative procedure must be used to find the minimum value that ensures optimality but in this case there is no such problem since the order-unit assignment is fixed. Thus, the number of event points to use is equal to the number of orders assigned to the machine plus one. Furthermore, each machine can be scheduled separately since the problems are in fact independent. Nevertheless, we will be solving all machines simultaneously as well as separately. The objective function was the (sum of the) total production time and the results are listed in Table 1.

Formulation CT4I exhibited a better computational performance, which is consistent with past studies (Castro et al. 2008). We then went on to address the full set of orders (73, corresponding to 16752 tons of paper), only to find out that CT4I failed to reach even a feasible solution. In contrast, CT3I successfully found a very good solution (0.3% relative gap) worth 57.60 days in about 4 hours. The same solution was returned after scheduling the machines separately, in a total of 5576 s (3562, 4.0 and 2010 in MP1, MP2 and MP3, for relative gaps of 0.5, 0 and 0.6%, respectively). Note that the computational statistics given are for the largest instance solved (MP1), the one with the most orders. CT4I got just to 57.77 days in significantly more time, a total of 17292 s. Thus, CT3I was preferred over CT4I.
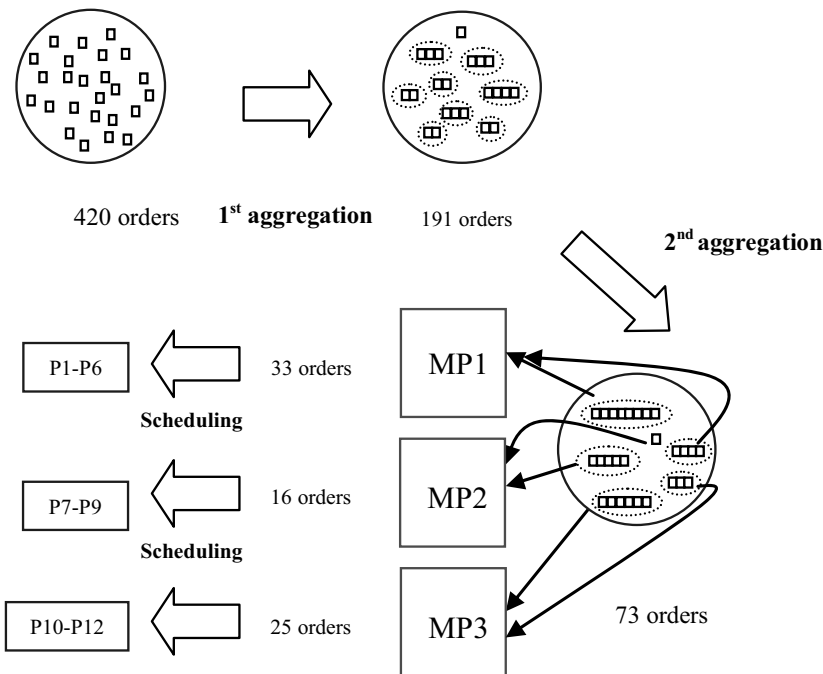
Figure 1. Two-stage solution strategy with order aggregation and fixed product-unit assignments

Table 1. Computational results for comparison CT3I vs. CT4I (CPLEX 10.2)

Model	# Orders	Machines	DV[a]	SV[b]	EQ[c]	OBJ (days)	CPUs
CT3I	36	Simultan.	1551	1595	1664	28.86	5042
CT4I	36	Simultan.	19212	20805	3085	28.86	3599
CT3I	73	Simultan.	2485	2567	5954	57.60	14800[d]
CT4I	73	Simultan.	59183	65032	6106	-	-
CT3I	73	Separately	1123	1158	1287	57.60	5576[d]
CT4I	73	Separately	34881	36071	1355	57.77	17292[d]

[a]Discrete variables; [b]Single variables; [c]Equations; [d]Maximum resource limit or computational time to reach relative optimality gap<1%

6. Comparison with a discrete-time approach

The schedule can also be derived with a discrete-time approach. One key difference is that in the second stage of the aggregation procedure described in section 4, the duration is now made as close as possible to a multiple of the interval length. The actual value, 0.25 days, results from a trade-off between total computational effort and data accuracy. Whenever the orders have combined duration lower than 0.25 days, more paper is produced to avoid the appearance of idle times, which is not tolerable in a continuous paper plant where some kind of paper is always being produced (excluding periods of maintenance. As a consequence, schedules from the discrete-time will produce 17.2% more paper (19428 vs. 16572 tons). The group of orders from the second aggregation procedure was also higher 124 vs. 73, which brings a natural increase in problem size.

The total production time was equal to 67 days, 16.3% higher than for continuous-time. All orders were produced before their due date and the maximum amount of paper in inventory was kept below the maximum capacity of the warehouse (6500 tons) even though no such constraint was in the model. In terms of the total number of changeovers, the continuous-time model was again the best performer (11 in MP1, 2 in MP2 and 4 in MP3 vs. 11, 5 and 8). More importantly, it reduced the number of changeovers in all three machines, when compared to the scheduling procedure at the plant for the same set of orders (12, 5, 5), indicating that a scheduling tool based on the continuous-time model can bring a higher level of productivity to the plant. Computationally, the higher complexity of the mixed-integer linear program (MILP) resulting from the discrete-time model, forced us to use a rolling-horizon scheme by further dividing the orders in two groups and scheduling them sequentially. Even with such scheme, the relative optimality gap was still significant for MP1 and MP3 after two hours of computational time (11.9 % and 2.9%) while MP2, featuring fewer orders, could be solved to optimality in 406 s.

7. From a sequencing to a scheduling problem

In order not to limit the study to the fixed order-unit assignment case decided by the plant, which significantly reduces the solution space and hence can severely compromise optimality, we released some products from this constraint. The choice took into consideration: (a) the number of orders; many would lead to a great increase in the computer effort;(b) frequency of occurrence of the maximum changeover time (30 minutes) between products allocated to the same machine. The outcome was to free P5 and P6, which represented 7 of the 33 orders allocated to MP1 by the aggregation procedure of the continuous-time model. We studied the effect of releasing either P5 or P6, and both products simultaneously. Naturally, the release process can only be tested if all machines are scheduled simultaneously and enough event points are added to the

time grids of MP2 and MP3 so that the orders can be relocated. Conceptually, we are making a switch from the solution of 3 sequencing problems to 1, more complex, scheduling problem. The results are given in Table 2.

In all three scenarios, the orders were reallocated to MP3, which was no surprise since this is the faster of the available machines. Its contribution to the total paper production increases from 62.1% to 66.7%. In terms of total production time, the reduction was significantly higher for P6 (4 orders) than P5 (3 orders) but was even greater for P5 and P6 (3.39%). It is particularly interesting to highlight that the P6 alternative is the best one in terms of makespan minimization (the maximum over all machines of the production time), leading to a makespan of 20.39 days. The computational effort has increased by over one order of magnitude. It is also interest to highlight that for the same relative optimality gap (1%) CPLEX 11.0 needs just 1269 s to solve the three separate problems, while CPLEX 10.2 required 14800 s. Note however, that the latter reached a slightly better solution (57.60 vs. 57.63).The optimal schedules are given in Figures 2-4.

8. Conclusions

This paper has addressed the optimal scheduling of a Portuguese paper plant. Given the high number of orders, in excess of four-hundred, a solution strategy was devised that consisted on order aggregation and fixed order-unit assignments followed by optimal sequencing of the three paper machines. Three alternative scheduling approaches were evaluated, two unit-specific continuous-time scheduling formulations for multistage multiproduct plants with sequence dependent changeovers and a discrete-time approach. The results showed that the continuous-time model handling changeovers implicitly was the best performer. Finally, we considered the case where a subset of the orders is no longer fixed to a particular unit to solve a more complex scheduling problem, where all machines are tackled simultaneously. Despite the order of magnitude increase in computational effort, we were able to decrease the total production time by 3.3% through the switching of production to the machine with the highest production rate.

References

P. Castro, I. Grossmann, A. Novais, 2006, Ind. Eng. Chem. Res. 45, 6210.
P. Castro, A. Novais, 2008, Ind. Eng. Chem. Res. 47, 6126.
Y. Liu, I. Karimi, 2007, Chem. Eng. Sci. 62, 1549.
C. Mendez, J. Cerdá, I. Grossmann, I. Harjunkoski, M. Fahl, 2006, Comput. Chem. Eng. 30, 913.
M. Shaik, C. Floudas, 2008, Comput. Chem. Eng. 32, 260.

Table 2. Computational results for comparison fixed vs. free P5/P6 assignments (CPLEX 11.0)

		Fixed	P5 free	P6 free	P5 & P6 free
Obj (days)	MP1	21.64	20.94 (-3.27%)	18.98 (-12.3%)	18.25 (-15.7%)
	MP2	16.80	16.80	16.80	16.80
	MP3	19.18	19.54 (+1.84%)	20.39 (+6.31%)	20.68 (+7.82%)
	Total	57.63	57.27 (-0.62%)	56.18 (-2.51%)	55.73 (-3.29%)
CPUs	Total	1269	15657	13325	20411
Production (ton)	MP1	5156	4989 (-3.23%)	4553 (-11.7%)	4386 (-14.9%)
	MP2	1192	1192	1192	1192
	MP3	10405	10571 (+1.60%)	11008 (+5.79%)	11174 (+7.40%)

Figure 2. Schedule for MP1 after releasing P5 & P6
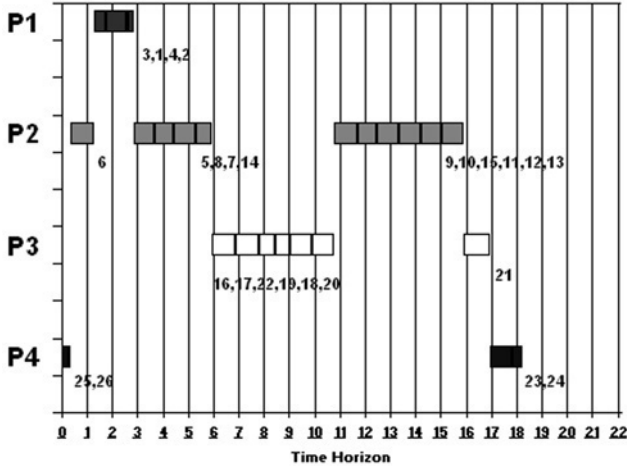
Figure 3. Schedule for MP2
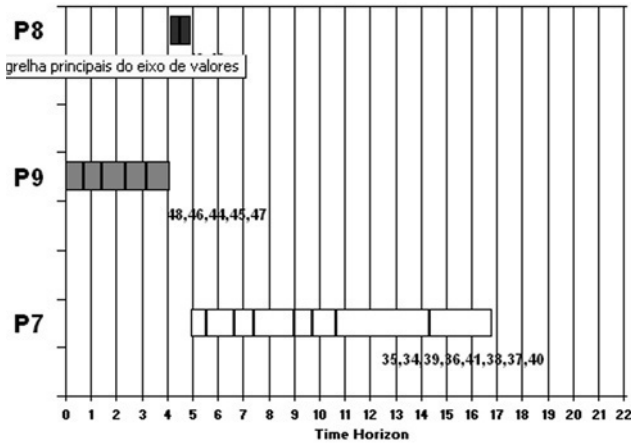
Figure 4. Schedule for MP3 after releasing P5 & P6
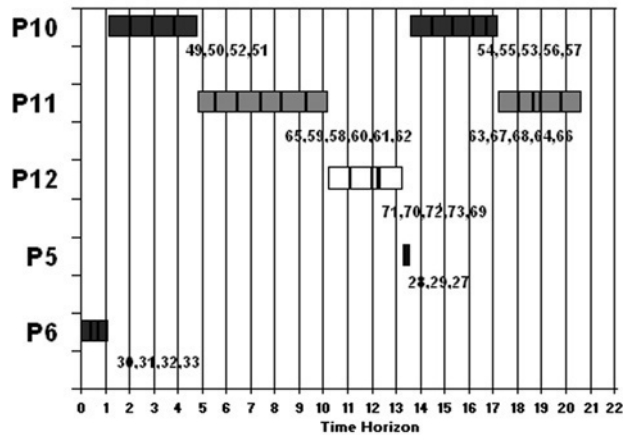
Entropic-Model-Based PI Controller

João T. Manzi, Antônio C.B. de Araújo, Romildo P. Brito and Heleno Bispo

Department of Chemical Engineering – Federal University of Campina Grande
Rua Aprígio Veloso n.882, Campina Grande-PB, CEP 58109-900, Brazil.

Abstract

Contributions of entropic modelling to the performance of reactive process control have been investigated. The modelling has been developed based on mass, energy and entropy balances and thermodynamics relations, resulting in a model for calculating the entropy production rate. Using the conventional optimization technique, a minimum for the entropy production rate has been found when a given relationship between the temperatures of the inlet stream and of the reaction is satisfied for a particular residence time in the reactor. A new class of nonlinear controller has been developed by means of introducing entropic models into classical PI algorithms designed by reference system synthesis. The results indicate that such a controller yields a superior performance when compared to classical feedback control strategies.

Keywords: PI controller, modeling, entropy, optimization.

1. Introduction

The use of models developed from a mechanistic (first-principles or phenomenological) point of view for the design, analysis, and control of chemical processes is a very powerful, underlying tool in process system engineering, and we are all used to considering such approaches as well-defined for the calculations. However, the thermodynamic treatment of such models are essentially based on the first law of thermodynamics, and do not usually consider the importance of the second law of thermodynamics as the limit factor on the direction of energy transformation.

Simultaneous use of the two laws of thermodynamics on the model building process has been the focus of important research studies. Recently, a procedure which incorporates direct minimization of entropy concepts as essential ingredients in the optimization of chemical process operations has been proposed (Manzi and Carrazzoni, 2008). The implications are such that operation at minimum entropy level can increase process yield while minimizing energy consumption. Furthermore, from a computational standpoint, the dimension of the optimization problem may also be reduced. Despite the advanced control strategies proposed for tackling industrial processes, the practical implementation of optimization policies, as in many other situations in the area of process engineering, usually relies upon the design of a decentralized control structure which makes use of the simplicity and recognized skills of PID controllers. However, it is well-known that these configurations do not yield satisfactory results in closed-loop for highly nonlinear processes when based on linear PID controllers.

Since entropic models can clearly reveal the optimal behavior and intrinsic relationships between the variables and parameters of the system, their introduction in PI algorithms designed by reference system synthesis, resulting in a nonlinear counterpart, can become a very attractive option to enhance closed-loop performance.

This article proposes the design of a class of nonlinear PI controllers for reactive systems using an entropic model in order to generate the control law.

2. Entropic modeling

Let a reactive system, depicted in Figure 1, be represented by the following generic reaction.

$$A_1 + A_2 + \cdots + A_i \rightarrow B_1 + B_2 + \cdots + B_i \qquad \Delta H < 0$$
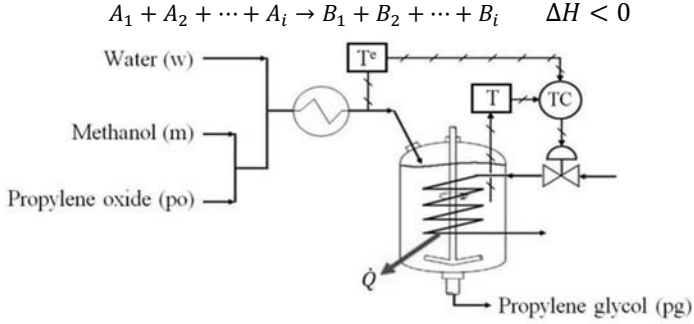
Figure 1: Diagram of a generic reactive system with a simplified feedback loop.

Consider a continuous stirred tank reactor (CSTR) as shown in Figure 1 in which a reactive system is presented consisting of A_i species reactants, resulting in B_i products. Such a reactive system has been mathematically described by means of the following equations representing the mass and energy balances respectively, besides kinetic considerations.

$$dn_{A_i}/dt = F^e C_{A_i}^e - F C_{A_i} - rV \tag{1}$$

$$dn_{B_i}/dt = -F C_{B_i} + rV \tag{2}$$

$$\left(\rho V c_p\right) dT/dt = -F^e \left(\sum C_i^e c_{p_i}\right)(T - T^e) + (-\Delta H_R)rV - \dot{Q} \tag{3}$$

Additionally, the entropy balance can be given by:

$$dS/dt = -\dot{Q}/T + F^e \rho^e s^e(T^e) - F\rho s(T) + \dot{\sigma} \tag{4}$$

where r and $\dot{Q}$ are respectively the reaction rate and the heat transferred from or to the process.

Since entropy is a function of temperature and of the number of moles of the constituent species of the system, then based on the concept of differential total and taking into account that ΔG_R represents the Gibbs free energy change of the reactive system, the following equation can represent the entropy generation rate for the system considered, as shown by Manzi and Carrazzoni (2008).

$$\dot{\sigma} = -F^e \left(\sum C_{A_i}^e c_{p_{A_i}}\right)[(T - T^e)/T + ln(T^e/T)] + rV(-\Delta G_R/T) \tag{5}$$

The so-called driving force $(-\Delta G_R/T)$ can be performed by using the classical Gibbs-Helmholtz relationship expressed by:

$$[\partial(\Delta G_R/T)/\partial T]_p = -\Delta H_R/T^2 \tag{6}$$

3. The synthesis of the PI controller

3.1. *Modelling for control purposes*

Consider the arrangement shown in Figure1. Since the rate of heat transferred $\dot{Q}$ derived from the energy balance applied to the cooling medium can be given by its linearized form around the point $(\dot{m}_c^e, T^e)$, then

$$\dot{Q}(\dot{m}_c, T) = \dot{m}_c(T^e - T^c)\left[c_{p_c} - c_{p_c}e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}} - (UA/\dot{m}_c^e)e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}}\right]$$
$$+UA(T^e - T^c)e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}} + \dot{m}_c^e c_{p_c}(T - T^e)\left[1 - e^{-\frac{UA}{\dot{m}_c c_{p_c}}}\right] \quad (7)$$

Using the Eq.(7) into the Eq.(3) and since the reactor behaves isothermally, the inlet temperature T^e can be chosen as the temperature of reference in the steady state, then, an equation, as a function of the deviation variables defined by $\bar{T} = T - T^e$ and $\bar{\bar{m}}_c = \dot{m}_c - \dot{m}_c^e$, results in:

$$(\rho V c_p)\,d\bar{T}/dt = -F^e\left(\sum C_i^e c_{p_i}\right)\bar{T} + (-\Delta H_R)rV - \bar{\bar{m}}_c(T^e - T^c)$$
$$\left[c_{p_c} - c_{p_c}e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}} - (UA/\dot{m}_c^e)e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}}\right] - \dot{m}_c^e c_{p_c}\bar{T}\left[1 - e^{-\frac{UA}{\dot{m}_c c_{p_c}}}\right] \quad (8)$$

3.2. *The reference system synthesis* (*RSS*)

The reference system synthesis, also known as the three steps synthesis (Bartusiak et al., 1989), can be outlined in the following stages, where *x, u, d* and *p* denote the state variable vector, the manipulated input, the disturbance and the model parameters respectively, *y* represents the controlled variable.

- *Stage 1*: deriving the model. $dy/dt = G_x f(x, u, d, p, t)$ where $G_x = \partial g/\partial x$ (9)

- *Stage 2*: specifying the reference system. $dy/dt|_{ref} = G_x f(x^*, x, p, l, t)$ (10)
where x^* and l denote the set point and the controller parameters respectively.

- *Stage 3*: minimizing the difference.
$$min[dy/dt - dy/dt|_{ref}] \rightarrow u = h(x^*, x, d, p, l, t) \quad (11)$$

Using the modelling previously developed, as well as taking into consideration desirable behaviour by the control system, such as to return toward its set point and to be free of offset, the following reference trajectory can be specified:

$$(\rho V c_p)d\bar{T}/dt|_{ref} = -F^e\left(\sum C_i^e c_{p_i}\right)\bar{T} + (-\Delta H_R)rV$$
$$-\left[k_c(\bar{T} - \bar{T}^{set}) + k_c/\tau_I \int(\bar{T} - \bar{T}^{set})dt\right](T^0 - T^c)\left[c_{p_c} - c_{p_c}e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}}\right.$$
$$\left.-(UA/\dot{m}_c^e)e^{-\frac{UA}{\dot{m}_c^e c_{p_c}}}\right] - \dot{m}_c^e c_{p_c}\bar{T}\left[1 - e^{-\frac{UA}{\dot{m}_c c_{p_c}}}\right] \quad (12)$$

Then, applying the reference system synthesis-based approach, the control algorithm can be found as:

$$\dot{m}_c = \dot{m}_c^e + k_c(\bar{T} - \bar{T}^{set}) + k_c/\tau_I \int(\bar{T} - \bar{T}^{set})dt \quad (13)$$

It is evident that Eq. (13) describes the PI control law. Since $\bar{T} = T - T^e$ can be related to the entropy production rate, the algorithm developed is in fact a controller based on the entropic concept.

4. Results and Discussion

To illustrate the analysis of such a control system when the process is submitted to a minimum entropy generation rate, consider, for simulation purposes, the production process of propylene glycol in a CSTR reactor reported by Manzi et al. (2009), and depicted in Figure 1. The reactive process occurs according to the reaction below, in which sulphuric acid has been used as a catalyst, while the maximum temperature should not exceed 324.8 K (Fogler, 1999). Table 1 provides the physical and chemical properties and the base steady state operating conditions applied to this system.

$$C_3H_6O + H_2O \rightarrow C_3H_8O_2 \qquad \Delta H < 0$$
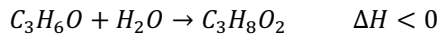
Table 1 – Operating conditions and parameters for the CSTR applied to the production of propylene glycol

Variable or	Value	Variable or	Value
F^e	2.567 L s^{-1}	$c_{p_{pg}}$	192.59 J mol^{-1} K^{-1}
τ	5,842.44 s	c_{p_m}	81.64 J mol^{-1} K^{-1}
C_{op}	2.12 mol L^{-1}	k_0	47.11x10^8 s^{-1}
T_r	298 K	E	75,320 J mol^{-1}
T^c	302.8 K	R	8.314 J mol^{-1} K^{-1}
Θ_w	18.65 -	U	567.83 J s^{-1} m^{-2} K^{-1}
Θ_m	1.67 -	A	35.7 m^2
$c_{p_{op}}$	146.54 J mol^{-1} K^{-1}	$\Delta H_{R_r}(298K)$	-84,589.11 J mol^{-1}
c_{p_w}	75.36 J mol^{-1} K^{-1}	$\Delta G_{R_r}(298K)$	-68,274.08 J mol^{-1}

4.1. Entropy production rate

According to Manzi et al. (2009), the necessary condition for the minimum entropy production rate requires $d\dot{\sigma}/dT$ and $d\dot{\sigma}/d\tau$ to be equal to zero. Thus the minimum value of $\dot{\sigma}$ only is achieved when the following relationships are satisfied:

$$(T^e - T) = (\tau k_0)/\left(\sum \Theta_i c_{p_i}\right)\left\{(E/R)e^{\left(-\frac{E}{RT}\right)}\Big/\left[\left(1 + \tau k_0 e^{\left(-\frac{E}{RT}\right)}\right)^2\right]\left(-\frac{\Delta G_r}{T_r} + \frac{\Delta H_r}{T_r}\right)\right.$$

$$+\Delta H_r\left[e^{\left(-\frac{E}{RT}\right)}\Big/\left[1 + \tau k_0 e^{\left(-\frac{E}{RT}\right)}\right] - (E/R)e^{\left(-\frac{E}{RT}\right)}\Big/\left(T\left(1 + \tau k_0 e^{\left(-\frac{E}{RT}\right)}\right)^2\right)\right]\right\} = 0 \quad (14)$$

$$\varphi(\tau) = \left[\frac{1}{\left(\tau k_0 e^{(-E/RT)}+1\right)^2}\right] = 0 \qquad\qquad\qquad (15)$$

where $\varphi(\tau)$ denotes the essential result of the differentiation procedure $(d\dot{\sigma}/d\tau)$.

Using the data provided by Table 1, the behavior of the temperature difference $(T^e - T)$ given by Eq. (14) can be depicted as shown in Figure 2, revealing that the intersection point between the $(T^e - T)$ curve with the temperature axis is the optimal

solution for Eq. (14), which indicates this to be $T^e = T$ the optimum operating temperature. As shown by Manzi et al. (2009), the optimized set of operating conditions derived from the entropy production rate yields the best value of conversion for the system, when compared to the classical modus operandi.
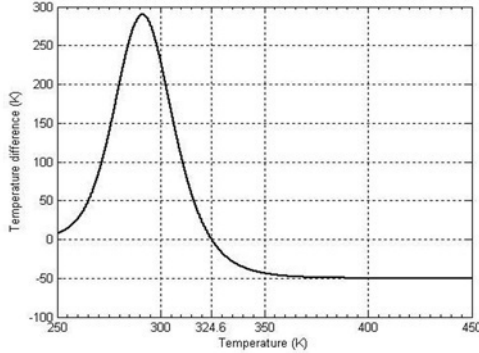
Figure 2: Temperature difference $(T^e - T)$ versus Temperature.

4.2. Entropic-model-based PI controller

The control structure, the so-called "entropic-model-based PI controller", results from Eq. (13) in which the signal of the error based on Eq. (14) is intrinsically related to the minimum entropy production rate. Since the temperature of the reactive system under the condition of the minimum entropy production rate requires $T^e = T$, then the input to the controller is given only by $\bar{T} = T - T^e$, that is, $\bar{T}^{set} = 0$, which results in the following nonlinear control law:

$$\dot{m}_c = \dot{m}_c^e + k_c \bar{T} + k_c/\tau_I \int \bar{T} dt \tag{16}$$

It is meaningful to observe that it is not necessary to specify the temperature desired for operating of the controller, because this temperature is implicitly determined by $\bar{T}^{set} = 0$, depicted in Figure 2, this being the unique solution for the data from table 1. It can be also observed that the search for the temperature desired follows a strategy which corresponds to the search for the minimum of the entropy production rate, and since, in practice, $T - T^e \geq 0$ is reflected straightforwardly in the settling time of the controller.

Figure 3 presents the comparative responses of the temperature of the reactive system when an entropic-model-based PI controller and a classical PI are used. In both cases, a disturbance has been introduced into the inlet temperature of the heat exchanger in addition to which the classical Ziegler-Nichols tuning procedure was applied with a fine adjustment based on trial and error in order to obtain the best set of tuning parameters. The results show that the entropic PI controller exhibits a much less pronounced oscillatory behavior, and has the lowest value for the integral of the absolute value of the error (IAE), with the response approaching its ultimate value asymptotically. It must also be emphasized that while the classical structure deals only with the controlled output, the entropic PI controller works under the condition of the minimum entropy production rate, and always yields the best performance for the whole reactive system.
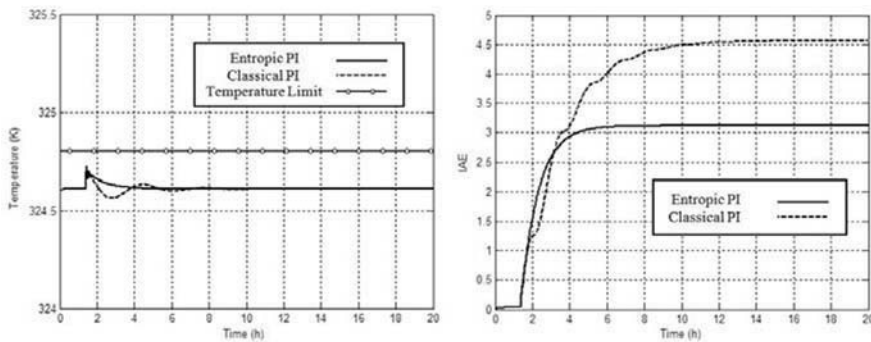
Figure 3: Temperature and IAE profiles from classical and entropic PI controllers when submitted to a disturbance in the inlet temperature of the heat exchanger from 290 to 295 K.

5. Conclusion

The entropic-model-based PI controller has been derived from relationships obtained by minimizing the entropy production rate, which allows the reactive system to reach the optimal conditions, in connection with the reference system synthesis. Since such optimal conditions have been introduced in the control structure, the control action moves the system towards the minimum entropy production rate, and satisfies the essential requirement $T = T^e$. Therefore, an advantage of such a control structure is due to the fact that it deals not only with the controlled output but above all with control of the whole process.

In addition, the entropic-model-based PI controller presents a shorter oscillating, faster response and the lowest IAE, and therefore reaches the desired value quickly and asymptotically. Thus, such results allow the conclusion to be reached that the entropic-model-basic PI presents a superior performance when compared with the conventional PI controller.

References

H. S. Fogler, 1999, Elements of chemical reaction engineering, Prentice Hall, New Jersey

J. Manzi, R. Viana and H. Bispo, 2009, Direct entropy minimization applied to the production of propylene glycol, Chemical Engineering and Processing: Process Intensification, 48, 1, 470-475

J. Manzi and E. Carrazzoni, 2008, Analysis and optimization of a CSTR by direct entropy minimization, Journal of Chemical Engineering of Japan, 41, 3, 194-199

R. D. Bartusiak, C. Georgakis and M. J. Reilly, 1989, Nonlinear feedforward/feedback control structures designed by reference system synthesis, 44, 9, 1837-1851

On-Line Fault Detection on a Pilot Flotation Column Using Linear PCA Models

Luis G. Bergh and Sebastián Acosta.

Chemical Engineering Department, Santa Maria University, Valparaiso, Chile

Abstract

On-line fault detection, for instrumentation and process operation, has become important part of industrial programs leading to improve process operation and therefore product quality over time. Multivariate statistical projection methods, such as Principal Component Analysis (PCA), have been proposed to effectively deal with these situations.

In this work, a pilot flotation column is operated under distributed control of froth depth, gas hold up and bias, to experimentally collect operation data at steady state, to build a PCA model. The basic control is implemented in a PLC, and all data is communicated to a PC network for displaying and further processing, under Intouch software. The column is operated in a hybrid form, for the air water system, while concentrate and tailing grades are obtained by on line predictions by using a static metallurgical model. A steady state on-line detector has been implemented on the PC to test when the collected data met the requirements to be used to build a PCA model. Several examples are discussed, detecting both particular instrumentation failures and abnormal operating conditions, and how using the system suggests the metallurgical objectives of the process are met again.

Keywords: Fault detection, modeling, projection methods, flotation columns, process control.

1. Introduction

In the last two decades the use of pneumatic flotation columns became wide-spread throughout the mineral processing industry of metallic, non-metallic and coal ores in the world. Columns out perform conventional mechanical cells in cleaning operations (better product grade) due to their particular froth operation (Finch and Dobby, 1990).

1.1 Flotation column control

The primary objectives are column recovery and concentrate grade, which represent the indices of process productivity and product quality. The on-line estimation of these indices usually requires a significant amount of work in maintenance and calibration of on-stream analyzers, in order to maintain good accuracy and high availability. Therefore, a common practice is to control secondary objectives, such as pH at the feed, froth depth, air flow rate and wash water flow rate. These are usually implemented as local controllers or under distributed control systems (DCS). Ideally, when primary objectives are measured, the control strategy is to change the set points of the

controllers under DCS, in order to achieve a good process performance. This is usually implemented in the form of expert systems (Bergh and Yianatos, 1993, 2003). On line analysers, tailings, feed flow rates and some other measurements are often incorporated into the system when a supervisory control strategy is implemented on top of a distributed control system.

1.2 Pilot flotation column

The pilot column was operated for the air-water (and frother) system coupled with an on-line steady state model to predict output stream grades. The convenience of the approach of combining on-line process measurements and models to empirically test strategies for process control, monitoring and diagnosis, was recently discussed by Bergh (2007). A flotation column phenomenological model was developed, following Finch and Dobby (1990). Following Figure 1, first the gas holdup, the bias rate and the kinetic constants for two mineralogical species are estimated from semi-empirical models, depending on operating variables such as feed flow rate, gas flow rate, wash water flow rate and froth depth. Then, dispersion number, residence times, froth and collection recovery are estimated. Finally concentrate and tailings grades are predicted. The semi-empirical model parameters were fitted using experimental data. More details can be found in Bergh et al. (1998).

Concentrate grade, tailings grade, recovery, bias rate, gas hold up, kinetic constants

Design parameters:
Diameter, height, geometry…

Process Metallurgic Simulator

Feed characteristics: species, density, solid percentage, particle size, grades, kinetic constants…

froth depth, air flow rate, water flow rate, feed flow rate
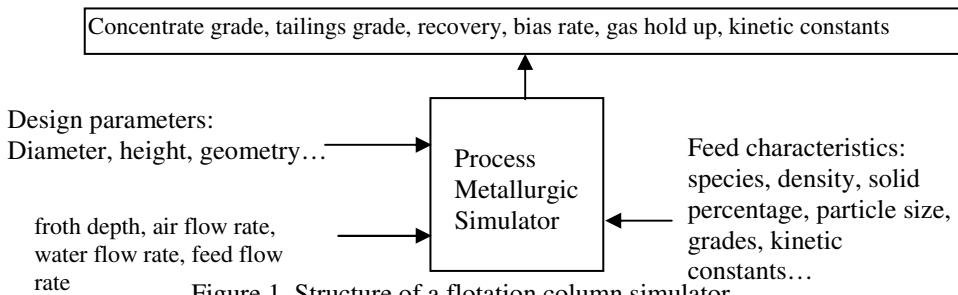
Figure 1. Structure of a flotation column simulator.

1.3 Pilot flotation column control

The column control is shown in Figure 2. There are three tertiary control loops: air flow rate, tailings flow rate and wash water flow rate. Feed flow rate is also measured and controlled. A hydrodynamic supervisory control is implemented to control gas hold up in cascade with air flow rate, froth depth in cascade with tailings flow rate and bias in cascade with wash water flow rate. All the operating variables are measured and communicated to a PLC, where the DCS has been implemented. All signals are communicated to a PC network, where the monitoring and hydrodynamic supervisory control is running under Intouch software. In the PC network, the steady state test is performed on-line. When the process reaches a steady state, the predicting model is solved on-line to estimate the concentrate and tailing grades. All data is displayed on PC screen.

2. PCA models

The key feature of PCA method is their ability to mathematically project high dimensional process and quality data into smaller dimensional, summary data sets via the development of linear models. The practical value of PCA modeling method is that

this technique allows for the systematic examination and interpretation of the model outputs. Examination of the model outputs can provide insight into the operation of an industrial process during monitoring and quality assurance activities. With PCA, the systematic interpretation of dominant patterns in the data and the isolation of the most important contributors to these patterns are possible. This allows the classification of data relationships according to normal and abnormal operation. Some of these numerous advantages of PCA method has over traditional monitoring and prediction technologies are: provision for data dimension reduction and robustness to highly correlated, noise and missing data (Kourti and MacGregor, 1995).

Fig. 2. Pilot flotation column control

The concept of a latent variable model is that the true dimension of a process is not defined by the number of measured variables, but by the underlying phenomena that drive the process. The latent variables themselves are modeled as mathematical combinations of the measured variables and describe directions of variation in the original data. A latent variable model can contain many fewer dimensions than the original data, it can provide a useful simplification of large data sets, and it can allow better interpretation of the measured data during analysis (MacGregor *et al*., 2007).

3. PCA model building

The original **X** matrix consisted of sixteen variables, shown in Table 1, and 2550 observations of steady state data. The experiments were designed to cover the maximum possible variation of the main independent variables, and were conducted under closed loop control. The data was processed using PLS_Toolbox from Eigenvalue Research.

A PCA model was built from 1800 sets of data corresponding to a normal condition. A model with 6 latent variables (scores) was found to explain at least 92 % of the variance in the centered and scaled pretreated data. The contribution of each operating and

quality variables (in the same order given in Table 1) to each PC is presented in Figure 3.

Table 1. Operating and quality variables considered.

N°	Variable	Tag	N°	Variable	Tag
1	Froth depth	z	9	Tailings superficial velocity	Jt
2	Gas hold up	E	10	Feed superficial velocity	Jf
3	Dp/cell low	LL	11	Wash water superficial velocity	Jw
4	Dp/cell high	LH	12	Cu recovery	R
5	Pressure to air control valve	PA	13	Concentrate Cu grade	CCG
6	Pressure to Tailings control valve	PT	14	Feed particle size d80	D
7	Bias superficial velocity	Jb	15	Feed Cu grade	FCG
8	Air superficial velocity	Jg	16	Feed solid percentage	S
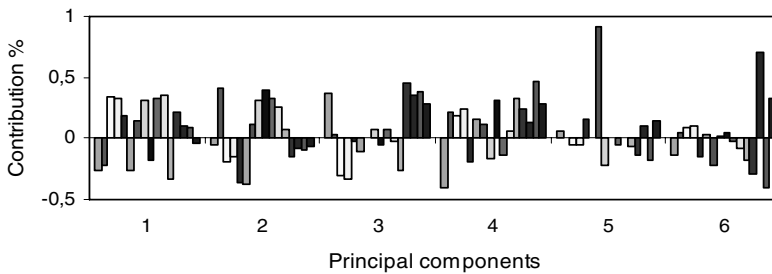
Fig. 3. Contribution of each sixteen operating and quality variables to each six PC

All variables were included in the model as a linear combination for each six scores. The first score showed the contribution of groups of correlated variables, while the last ones mostly represented the bias and the feed characteristics. For monitoring the process the Hotelling T^2 limit was found to be 12.6, while the Q residuals was 3.81.

On-line tests were implemented based on the following criteria:

(i)   Normal operation if the new set of data satisfies the Q and T^2 test.
(ii)   Abnormal operation if the T^2 test is failed. If the Q test is passed then the model is adequately representing the process. If the Q test is also failed then either the model is no longer appropriate or a measurement problem occurred.
(iii) Measurement problem or PCA model representation problem if only T^2 test is satisfied.

In this way a diagnosis of the operation can be accomplished for steady state data. Furthermore, the residuals are informative of the principal process variables affecting the abnormal situation.

4. Experimental results

Experiments were carried on to test when the process is out of control and an abnormal operating condition is met. Two results are presented: when the process is at steady state and during the transient period. One example is shown in Figure 4, where the T^2 and Q test has been followed for over 600 samples, taken every 5 (s).

One can see that most of the time the Q test is satisfied, while T^2 test is failed at intervals 130-200, 300-430 and 480-560. On these same periods, the concentrate grade is too low and recovery is high or concentrate grade is too high and recovery is low, then an abnormal operation has been detected. To identify which variables are causing this, the individual contribution to the T^2 residuals, for sample 512, are shown in Figure 5. One can see that the main contribution were the froth depth and the high and low dp/cells. All variables consistently showed that the problem is due to a low froth depth, causing high recovery and low concentrate grade. Figure 6 shows the froth depth changes during the whole period. If the froth depth were change from 50 to 100 cm, as is shown at sample 600, the column operation is driven back to a normal condition, as can be seen from the previous figures.

Fig. 5. Contributions to T^2

Fig. 6. Froth depth period

Fig. 4. Operating condition test
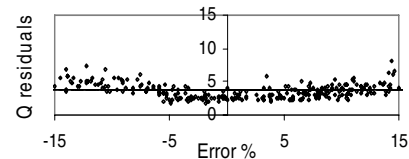
Fig. 7. Failure on concentrate grade

When only the Q residuals test fails, the device measuring the isolated variable must be recalibrated o replaced. Several tests were carried on to find the sensibility of the monitoring test to the extension of the fail, measured in percentage of error. Errors less than 5% on pressure to control valves, 7% on Dp/cells, 15% on flow meters and 10% on virtual measurements of concentrate grade were detected. These error limits were found for a large number of different operating conditions. One example is shown in Figure 7 for the virtual measurement of copper concentrate grade.

The same PCA model was used to test abnormal operation either because of decision based on failed sensors or process variable deviations. The PCA model relays on the selected data. If the data collected represents a narrow band of operation around the targets, it may be expected that abnormal conditions, as a result of a combination of

process variable deviations, will be easily detected. A model built on such selected data will be less useful to identify measurements problems. The model used in this work was based on data corresponding to a wide operation zone, favouring the detection of sensor failures. A best approach to be tested is the use of different PCA models, based on different data, for each purpose.

Another surprising aspect was that even when the process was in a transient, and the grades predictions were not entirely valid, the on-line application of the T^2 and Q test provided information about when an abnormal operation was in progress. Further research is needed to asses this fact.

5. Conclusions

Flotation column control quality is strongly depending on the accuracy of measurements and estimations. The flotation process is complex and it is a real challenge to decide which variables are to be changed in order to drive back the process to a normal operation.

The application of multivariate statistical methods, and particularly PCA, is a powerful tool to build linear models containing the essential of the process phenomena with the minimum number of latent variables. The application of PCA models to monitoring flotation columns has been demonstrated. These PCA models can be effectively used as part of a supervisory control strategy, specially when control decisions are infrequently made.

References

Bergh L.G. and J.B. Yianatos (1993). Control Alternatives for Flotation Columns. *Minerals Engineering*, **6**, Nº 6, 631-642.
Bergh L.G. (2007). Combining on-line process measurements and models to empirically test strategies for process monitoring, diagnosis and control, *Proceedings 12th IFAC Symposium MMM 07, Quebec City, August 21-23,* 193-198.
Bergh L.G. and J.B. Yianatos (2003). Flotation Column Automation: state of the art. *Control Engineering Practice*, **11**, 67-72.
Bergh L.G., J.B. Yianatos and C. Leiva (1998). Fuzzy Supervisory of Flotation Columns. *Minerals Engineering*, **11**, N° 8, 739-748.
Finch J.A. and G.S. Dobby (1990). *Column Flotation*. Pergamon Press.
Kourti T. and J.F. MacGregor, (1995). Process Analysis Monitoring and Diagnosis using Multivariate Projection Methods – A Tutorial, *Chemometrics and Intelligent Laboratory Systems*, **28**, 3-21.
MacGregor J.F., T. Kourti, J. Liu, J. Bradley, K. Dunn, H. Yu (2007). Multivariate methods for the analysis of databases. process monitoring, and control in the material processing industries, *Proceedings 12th IFAC Symposium MMM 07, Quebec City, August 21-23,* 193-198.

1443

Measurements of Air Quality Using Lidar System

Juliana Steffens[a], Eduardo Landulfo[b], Roberto Guardani[a], Cláudio N. Oller[a], Andréia Moreira[c]

[a] *Universidade de São Paulo, Escola Politécnica, Departamento de Engenharia Química, Av. Luciano Gualberto Trav. 3 N. 380, Butantã, 05508-900, São Paulo - SP, Brazil.*
[b] *IPEN-Instituto de Pesquisas Energéticas e Nucleares, Avenida Prof. Lineu Prestes 2242,Cidade Universitária, 05508-000, São Paulo, Brazil*
[c] *PETROBRÁS, Petróleo Brasileiro S.A., Ilha do Fundão, Rio de Janeiro, Brazil*

Abstract

Air pollution is a widely recognized hazard to human health. In industrial cities the emission of toxic gases and particulate matter create hazardous public health situations. Is the case of the industrial complex of Cubatão, state of São Paulo that is one of the largest petrochemical and industrial in Brazil that has been subject of severe damage caused by massive emissions of pollutants, as a result of the progressive industrialization in the area. Therefore it is necessary to monitory the area to be able to control and to prevent ambient problems. In a partnership with the University of São Paulo (USP) the Brazilian oil company PETROBRAS has started off an Environmental Research Center - CEPEMA- located in the industrial site, in which the development of fieldwork will be carried out. The current joint R&D project focus on the development of real time acquisition system, together with automated multicomponent chemical analysis. Additionally fugitive emissions from oil processing and storage sites will be measured, together with the main greenhouse gases (CO_2, CH_4) and aerosols. Our first effort is to assess the potential chemical species coming out of an oil refinery site and to use Raman technique LIDAR for detecting and quantifying the particular gas. Raman lidar techniques have been demonstrated which provide most valuable descriptions of the evolution of air pollution events. The vibrational and rotational Raman LIDAR signals provide simultaneous profiles of meteorological data, ozone and measurements of airborne particulate matter.

Keywords: Raman LIDAR, air quality, fugitive gas.

1. Introduction

Methane plays an important role in the Earth's atmospheric chemistry and radioactive balance. It is the most abundant hydrocarbon in the atmosphere, and the second most important greenhouse gas after CO_2. It has a relatively short lifetime (8-12 years), but its strong IR absorption band at 7,66 μm, where water and CO_2 absorb weakly, makes methane an effective contributor to the radiative forcing with a global warming potential of 21 over a 100 year period (IPCC, 2001). Another important role in the atmosphere is the atmospheric aerosol, because is a complex mixture of particles from a large number of discrete sources (Ondov and Wexler, 1998). Its size distribution, composition, morphology, and source strenghts can vary significantly with meteorology, location, and time. This is especially true for urban atmospheres wherein emissions from

concentrated industrial and dispersed sources lead to urban pollutant excesses above rural background concentrations (IPCC, 1996).

Today, most industrial plants are obligated to measure their own emissions. Accurate measurements of the concentration and the flux of different gases are important, not only from an environmental point of view, but also for economical reasons, since a part of taxation may be dependent on the emission values. Different types of in situ instruments are frequently used to monitor the main emission sources, but external optical methods are becoming increasingly important. The advantages with optical methods are online evaluation of the measurement, no contamination of samples, measurement a long distances and the possibility of measuring large areas (Sigrist, 1994).

Certain industrial activities are of key influence on the air quality and nowadays have triggered the awareness over long term atmospheric changes as well. Their monitoring became not only a problem of mitigating human health impact due to air pollution but also a matter of climate change. Among the industries with influence on the air quality the petrochemical sector plays one of the most important roles since the economical growth has augmented the diversity of products and storage capacity.

Following this increase in activity the need for monitoring systems in these sites and their neighboring areas have also become tasks of great concern, especially because data acquisition and analysis in practically real time have turned into a sine qua non feature in such systems. Therefore the technique and equipments to be employed have to be capable of detecting in real time the emission of aerosols (particulate matter), polluting air gases, such as NOx, Ozone, Sulphur Compounds and fugitive gases in industrial and storage sites(namely Volatile Organic Compounds - VOCs). Recent advances of tunable all solid state laser systems and in the Lidar technique (Light Detection and Ranging) opened new perspectives in the 3D-analysis of atmospheric pollution dynamics. 3D mappings of concentrations of pollutants have been obtained, allowing a direct access to the physical and chemical dynamics of air pollution (Frejafon *et al*, 1998). Besides their detection, these species have to be tracked along neighboring areas by the use of dispersion models. Given the extension and quantities involved, meteorological sensors and even satellites have to be employed. Based on the situation described, the use of LIDAR-based technique is proposed in the region of Cubatão, Brazil, one of the largest petrochemical and industrial sites in Brazil, where petrochemical activities exist since about 50 years.

In a partnership with the University of São Paulo (USP) the Brazilian oil company PETROBRAS has created CEPEMA, an Environmental Research Center located in the industrial site, where the development of the monitoring technique is supposed to take place. It is intended to develop a system capable of real time acquisition and by the use of automated multicomponent chemical analysis techniques, this system should be able to identify fugitive gases from oil processing and storage sites, and also the main greenhouse gases, CO_2 and CH_4; and aerosols. The use of ancillary equipment should be carried out for validation and calibration purposes. The system should operate with a Fourth-Harmonic-Generated Nd:Yag UV laser source (266 nm) and detect the Raman scattered lines of the species mentioned above. Lower harmonics (355 nm, 532 nm and 1064 nm) should be used for the aerosol backscattered signal, and also capable of particle size determination. This paper should cover the specific needs for a air monitoring spectrocopy system in operation in situ in a refinery site.

2. Oil Refinery Gas Inventory

The air environment in an petrochemichal refinery is rich in hydrocarbon compounds which have rotationalvibrational transitions in the mid-IR wavelength region. Of the most popular carbon-containing compounds present in such an environment we highlight the class of alkanes or paraffins, such as Methane CH_4, Ethane C_2H_6, Hethane C_6H_{14} and Hexane C_4H_{10}, the class of alkenes, such as Ethene or ethylene C_2H_4 and Propene or propylene C_3H_6 , followed by the Alkynes with Acetylene C_2H_2, Aromatics Benzene C_6H_6 and Toluene C_7H_8 and finally the Aldehydes, Formalaldeyde *HCHO* and Acetaldehyde *CH_3CHO*. Its is estimated that globally around 17.5 Tg of Nonmethane Volatile Organic Compounds is emitted to the atmosphere by the fuel production/distribution industry, from that about 30% is due the Oil refining branch (Seinfeld and Pandis, (1997).

3.Raman Lidar

3.1- Lidar remote sensing

The basis for lidar remote sensing lies in the interaction of light with gas molecules and particulate matter in suspension in the atmosphere (aerosols). More particularly, a lidar uses a laser (emitter) to send a pulse of light into the atmosphere and a telescope (receiver) to measure the intensity scattered back (backscattered) to the lidar. By measuring the scattering and attenuation experienced by the incident pulse of light, one can investigate the properties of the scatterers (concentration of gaseous species, aerosol distribution and optical properties, cloud height) located in the atmosphere. The light scattered back to the detector comes from various distances, or ranges, with respect to the lidar. Because the light takes longer to return to the receiver from targets located farther away, the time delay of the return is converted into a distance (range) between the scatterers and the lidar, since the speed of light is a well-known quantity. By pointing the laser beam in various directions and at various angles with respect to the ground surface (scanning), a ground-based lidar system can gather information about the three-dimensional distribution of aerosols in the atmosphere.

3.2 – Raman technique

Raman scattering is one of the processes that occurs when optical radiation is scattered from the molecules of the atmosphere. It is most useful because the vibrational Raman scattering provides distinct wavelength shifts for species specific vibrational energy states of the molecules and rotational Raman scattering provides a signal with a wavelength shift that depends directly upon the atmospheric temperature (Philbrick, 1994).
Figure 1(a) shows a diagram of the vibrational and rotational energy levels that are associated with Raman scatter. When a photon scatters from a molecule, the redistribution of the charge cloud results in a virtual energy state. Most of the atmospheric molecules reside in the ground vibrational level because the vibrational excitation corresponds to relatively large energy transitions (tenths of eV), for simple molecules like nitrogen and oxygen, compared to the thermalenergy available. After the scattering occurs, most of the events result in the return of the molecule to the ground state and the emitted photon has the energy of the initialphoton plus/minus the random thermal velocity of the molecule, that is the Doppler broadening. A small fraction of the

transitions (order of 0.1%) result in giving part of the photon energy to the molecule, and ending in the first vibrational level (a Stokes transition). The emitted photon energy is decreased by exactly the energy of the vibrational quanta for that mole cule. For the small fraction of molecules existing in the vibrational excited level, the unlikely anti-Stokes transition is possible. The relative sensitivity of the scattering from the vibrational and rotational states is indicated by the scattering cross-section values for scattering by a frequency doubled Nd:Yag laser at 532 nm shown in Figure 1(b). The wavelengths of vibrational Raman back scatter signals from the molecules of the water vapor and molecular nitrogen are widely separated from the exciting laser radiation and can be easily isolated for measurement using modern filter technology and sensitive photon counting detectors (Balsiger, and Philbrick, 1996).
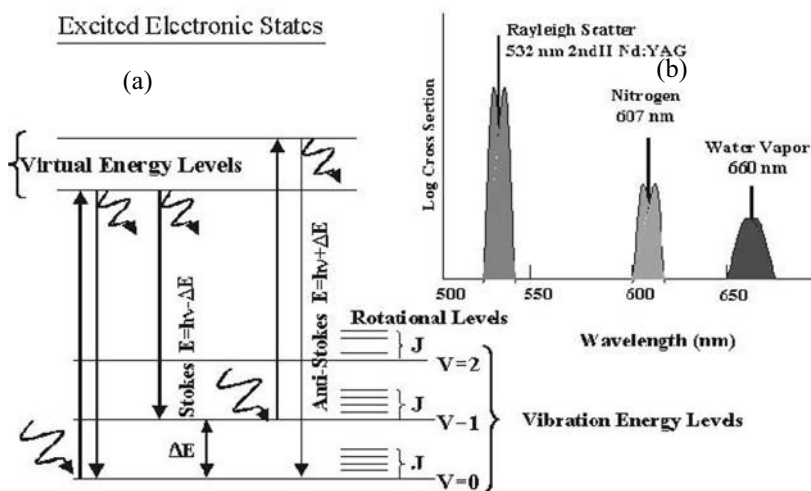
Figure 1 (a) The energy diagram of a molecule illustrates that the scattering of a photon raises the molecule to a virtual level which normally decays to ground (V=0) emitting a photon of the same energy as the incident energy, only broadened by thermal Doppler velocity. In a small fraction of cases, the return is Raman shifted to the first vibrational level (V=1), a Stokes shift. The relatively large vibrational energy (ΔE) compared with thermalenergy makes the Anti-Stokes vibrational transition unlikely,however, the rotational states (J-levels) are populated by thermal excitation, (b) The relative intensities of the Stokes vibrational Raman shifts of oxygen, nitrogen and water vapor are indicated for illumination of atmospheric molecules with the 532 nm laser. The expected line widths and the relative rotational states are indicated (Philbrick, 1994).

The ratio of rotational Raman signals at 528 nm and 530 nm provides a measurement which is sensitive to atmospheric temperature (Balsiger *et al.*, 1996). All of the molecules of the lower atmosphere are distributed in the rotational states according the temperature. By measuring the ratio scattered signals at two wavelengths in this distribution, the temperature can be directly measured. In order to push the lidar measurement capability into the daylight conditions, we have used the "solar blind" region of the spectrum between 260 and 300 nm. The "solar blind" region is darkened by the stratospheric ozone absorption of ultraviolet radiation. Night time measurements are made using the 660nm/607nm (H_2O/N_2) signal ratio from the doubled Nd:Yag laser radiation at 532 nm. Daylight measurements are obtained using the 295nm/284nm

(H_2O/N_2) ratio from the quadruple Nd:YAGlaser radiation at 266 nm. A small correction for the tropospheric ozone must be applied. That correction can be obtained from the ratio of the O_2/N_2 signals 278nm/284nm, and from this analysis the ozone profile in the lower troposphere is also obtained (Esposito and Philbrick, 1998) The Raman techniques, which use ratios of the signals for measurements of water vapor and temperature, have the major advantage of removing essentially all of uncertainties, such as any requirement for knowledge of the absolute sensitivity and non-linear factors caused by aerosol and cloud scattering (Philbrick, 1998). Optical extinction is measured using the gradient of the measured molecular profile compared with that expected for the density gradient. Since the Raman signal is only scattered from the molecular component of the scattering volume, and difference in the gradient of the signal from that expected due to loss frommolecular scattering and absorption can be used to calculate the aerosol extinction (O'Brien *at al*.,1996).

4.Lidar system design

This system should be deployed at CEPEMA, Centro de Pesquisa do Meio Ambiente, Cubatão, in a neighboring region to a PETROBRAS refinery station. This system should be built not only to monitor the air quality but also to diagnose the refining process performance by inspecting the particle size distribution coming out of the chimneys. The telescope employed should be able to scan from 0^φ to 90^φ azimuthally and thus be capable of full spatial coverage.

The dates will be analyzed by Statistical technical to interpreted results.

At Figure 2 we present one Schematic diagram of Raman Lidar that should use in the experiments.

1.Nd-Yag laser
2.Scanning mirror
3.Telescope
4.Photomultiplayer
5.Photon counting electronics
6.Data acquisition computer
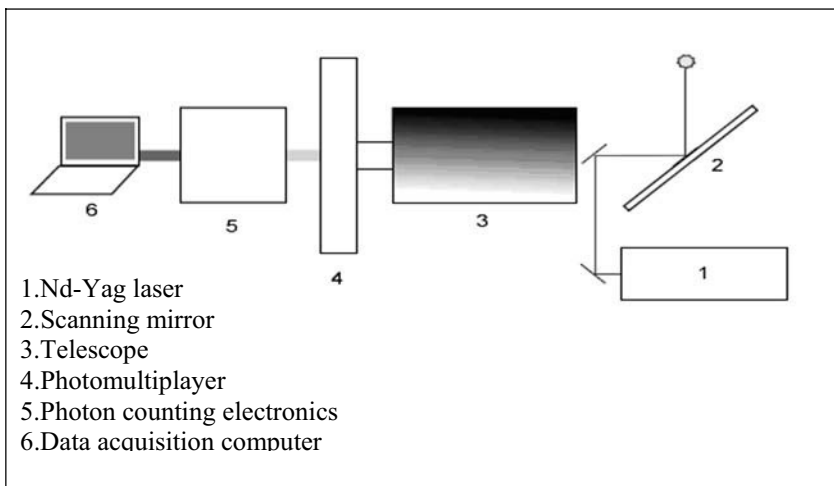
Figure 2- Schematic diagram of Raman Lidar.

5.Conclusion

Lidar technology has matured over the last decade to an extent that many applications are becoming routine. Meteorological parameters like the wind and temperature, as well as trace gas concentrations, can be measured in short time periods with high spatial resolution in three dimensions, often at large distances. Plumes containing air pollutants like hydrocarbons , VOCs and others can be measured by lidar at strategic sites.

The current joint R&D project focus on the development of real time acquisition system, together with automated multicomponent chemical analysis. Additionally fugitive emissions from oil processing and storage sites will be measured, together with the main greenhouse gases (CO_2, CH_4) and aerosols. Our first effort is to assess the potential chemical species coming out of an oil refinery site and to verify if Raman Lidar is efficient in detecting and quantifying the particular gas.

References

Intergovernmental Panel on Climate Change, 2001. Climate change 2001. In: Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., Van Der Linden, P.J., Dai, X., Maskell, K., Johnson, C.A. *The Scientistic Basis*. Cambridge University press, UK.

Intergovernmental Panel on Climate Change, 1996. Climate change 1996. The science climate chenge. Cambridge University press, New York.

M.W.Sigrist, 1994. Air monitoring by spectroscopic techniques, *Chemical analysis*, Vol, 127, New York.

J.M. Ondov and A.S.Wexler, 1998. Where do particulate toxins residue? An improved paradigm for the structure and dynamics of the urban mid-atlantic aerosol, *Environmental Science Technology*, Vol. 32, p. 2547-2555.

E.Frejafon, J. Kasparian, P. Rambaldi, B., Vezin, V, Boutou, J. Yu, M. Ulbricht, D. Weibauer, B. Ottobrini, E. de Saeger, B. Kramer, T. Leisner, P. Rairoux, L. Woste, J.P., Wolf. (1998) Laser applications for atmospheric pollution monitorin. *The European Physical Journal D*, Vol.4, p. 231-238.

C.R.Philbrick, (1994),Raman Lidar Measurements of Atmospheric Properties, *Atmospheric Propagation and Remote Sensing III*, SPIE Vol. 2222, p.922-931.

F. Balsiger, and C. R. Philbrick, (1996), Comparison of Lidar Water Vapor Measurements Using Raman Scatter at 266nm and 532 nm, *Applications of Lidar to Current Atmos. Topics*, SPIE Proc. Vol. 2833,p. 231-240.

F. Balsiger, P. A. T. Haris and C. R. Philbrick, (1996), Lower-tropospheric Temperature Measurements Using a Rotational Raman Lidar, *Optical Instruments for Weather Forecasting*, SPIE Proc. Vol. 2832, 53-60.

S. R. Esposito and C.R. Philbrick, (1998), Raman/DIAL Technique for Ozone Measurements, *Proceeding of Nineteenth International Laser Radar Conference*, NASA/CP-1998-207671/PT1, p.407-410, 1998.

C.R. Philbrick (1998), Raman Lidar Capability to Measure Tropospheric Properties,*Proceeding of Nineteenth International Laser Radar Conference*, NASA/CP-1998-207671/PT1, p. 289-292.

M.D.,O'Brien, T. D. Stevens and C. R. Philbrick, (1996), Optical Extinction from Raman Lidar Measurements, *Optical Instruments for Weather Forecasting*, SPIE Proceedings Vol. 2832, p. 45-52.

J. Seinfeld and S. Pandis, (1997), *Atmospheric Chemistry and Physics*, John Wiley and Sons, Inc.

Energy Efficiency in an Industrial Wet Cooling Tower Through Improved Control

Celso A. X. Marques[a,1]; Cristiano H. Fontes[a,2]; Marcelo Embiruçu[a,3]; Ricardo A. Kalid[a,4].

Polytechnique School of Federal University of Bahia - Industrial Engineering Programme (PEI), Rua Aristides Novis, 2, Federação, Salvador, BA, Brazil. CEP-40210-630. TEL: (55) (71) 3283-9505 FAX: + (55) (71) 3283-9800.

Abstract

Cooling towers are too much used in process plants in order to allow heat removal from the process to the atmosphere. These pieces of equipment are designed for achieve the maximum performance at design conditions. However, under other conditions such as lower heat load and lower wet bulb temperature, which frequently occur during plant operation, there is an excess of cooling capacity which is not used. This paper presents an open-loop analysis of an industrial wet cooling tower and proposes a closed-loop strategy in order to control two indexes of performance (efficiency and effectiveness). The closed-loop results suggest the possibility of a sensible improvement in the eco-efficiency of this piece of equipment thorough an expressive reduction in its energy consumption, without dropping of performance regarding the satisfaction of the thermal requirements of the process.

Keywords: Cooling towers, eco-efficiency, dynamic simulation, split-range control.

1. Introduction

Cooling towers have been neglected as an important aspect regarding to the eco-efficiency in industrial processes. The analysis and implementation of optimization and/or control techniques in cooling towers represent a potential source of improvement of energy efficiency and reduction in water consumption. Some works consider the steady state thermal performance of cooling towers and also present some inconsistencies in the results predicted by the Merkel's model (Nahavandi and Oellinger, 1977; Sutherland, 1983; Khan and Zubair, 2001). Different approaches have been proposed in modeling, optimization and performance evaluation of cooling towers. Majumdar et al. (1983) presented a two-dimensional model based on mass, energy and momentum balances considering two case studies, namely, mechanical draft and natural draft. Dreyer and Erens (1996) proposed a one-dimensional model based on aerodynamic, hydrodynamic and mass and heat transfers for the evaluation and design of splash packing geometries. Dessouky et al. (1997) presented a modified expression for calculating the effectiveness and the NTU taking into account the resistance of heat transfer in the water film. Söylemez (2004) used the criterion of effectiveness for the thermo-hydraulic performance optimization and analysis of forced draft counter flow cooling towers. Hosoz et al. (2006) applied a neural network for the prediction of the

performance of a pilot tower in terms of the following parameters: heat loss rate, water evaporation rate, and outlet water temperature and outlet conditions of the air stream.

2. Dynamic modeling, simulation and parameter estimation

The phenomenological model adopted to describe the dynamic behavior of the counter flow wet cooling tower was formulated through energy and mass balances under unsteady state conditions considering the volume element of packing (Younis et al., 1987). The equations comprise a one-dimensional model with distributed parameters and the line method was used in the dynamic systems simulation. The derivatives of position were approximated through finite differences and the time integration was carried out using the Klopfenstein's method (Ramirez, 1997). The fill packing volume was divided into 10 sections, leading to a differential algebraic system with 30 ordinary differential equations and 10 algebraic equations. An additional equation related to the energy balance of the heat exchanger which represents the process heat load was also considered in the model. Normal operation data (open-loop and unsteady state conditions) collected from a commercial cooling tower operating in a cellulose production unit located in Brazil were used to estimate the model parameters. The design specifications of this equipment are presented in Table 1.

Table 1 – Design specifications of the commercial cooling tower

Thermal capacity	132 GJ/h
Tower dimension	22.3 m × 11.95 m × 6.7 m
Fill packing height	1.34 m
Water circulation flow rate	2100 m^3/h
Inlet water temperature	45.0 °C
Outlet water temperature	30.0 °C
Number of cells	2
Air flow rate (each fan)	17124 m^3/min
Nominal fan power	102 kW

2.1. Open-loop analysis

A step response analysis was carried out through a change from the initial steady state condition in each of the following inputs: water flow, air flow, wet bulb temperature and heat load. The water and air flows represent possible manipulated variables and the other variables are disturbances produced by the environment (wet bulb temperature) and by the process (thermal demand). Two index of performance are investigated, namely, efficiency (Fisenko and Petruchik, 2004) and effectiveness (Khan and Zubair, 2001). The efficiency (η) is associated with the distance of the outlet water temperature from the wet bulb temperature, according to the equation 1:

$$\eta = \frac{T_{w,1} - T_{w,2}}{T_{w,1} - T_{wb}} \tag{1}$$

where $T_{w,1}$ and $T_{w,2}$ are the water temperatures in the inlet and outlet streams, respectively, and T_{wb} is the wet bulb temperature. The effectiveness (ε) is related to the distance of the outlet air humidity from its saturation condition evaluated at inlet water temperature, according to the equation 2:

$$\varepsilon = \frac{i_2 - i_1}{i_{asw,2} - i_1} \tag{2}$$

where i_1 and i_2 are the specific enthalpies of the air in the inlet and outlet streams, respectively, and $i_{asw,2}$ is the specific enthalpy of the saturated air, in the outlet stream, evaluated at the inlet water temperature.

As a result of the step response ($\pm10\%$) in the water flow rate, a first order behavior with some differences between the gains in positive and negative step responses was observed for the effectiveness. The same disturbance applied to the air flow rate showed an inverse response in the efficiency. The efficiency presented over-damped behavior for both process heat load and wet bulb temperature disturbances. The summary of the open-loop analysis is showed on Table 2 that presents the gain signals of each input-output pair. In all cases, efficiency and effectiveness have opposite signs for each input.

Table 2 – Gain signals: open-loop analysis

	Efficiency	Effectiveness	Outlet water temperature
Air flow	+	-	-
Water flow	-	+	+
Process heat load	+	-	+
Wet bulb temperature	+	-	+

3. Cooling tower control

Cooling towers are designed taking into consideration extreme conditions of operation, namely, the highest wet bulb temperature and the maximum heat load. In general, this results in oversized cooling capacity of the installations. Therefore, in the absence of any regulatory control it is impossible to achieve satisfactory performance and high energy consumption usually results. Despite the importance of appropriate control of cooling tower operation to save energy, few works have been reported in the literature. Dijk et al. (1985) considered the use of the water flow to keep the hot water temperature at its design value together with the air flow in order to control the cold water temperature. Moreover, the set-point of the cold water stream is established according to the wet bulb temperature trajectory. Fonstad (1988) considers that only the air flow can be frequently manipulated and states that a PID controller could be used to keep the hot water temperature at the desired set-point. Chen and Liang (2005) applied fuzzy control to provide a reduction in energy consumption in frigorific equipment's cooling system. According to Stout and Leach (2002) there is less potential for energy saving in cross flow rather than counter flow cooling towers through the manipulation of the tower fan. Fisenko and Petruchik (2004) showed that it is possible to reduce the energy consumption of the fan while both air humidity and local air temperature decrease.

Here, a split-range configuration is adopted and the control strategy comprises the control of a single process output by coordinating the actions of several manipulated variables. Based on the velocity algorithm of the digital PID controller, the controller output signal is:

$$\Delta c(k) = Kc \cdot \Delta e(k) + \frac{Ts \cdot Kc}{\tau_I} \cdot e(k) + \frac{Kc \cdot \tau_D}{Ts} \cdot \Delta^2 e(k) \tag{3}$$

where $\Delta c(k)$ is the variation in the controller output signal at instant k, T_S is the sample period, $\Delta e(k)$ and $\Delta^2 e(k)$ are the error variation and the variation of the error

variation, respectively. Parameters Kc, τ_I and τ_D are the tuning parameters (controller gain, integral time constant and derivative time constant). The current value of each manipulated variable can be obtained directly through the variation $(\Delta c(k))$ of the control output signal sent to each final control element. The signal sent to the water pump variable-frequency drive (VFD) is given by:

$$\Delta c_1(k) = -\Delta L \tag{4}$$

and the signal sent to the fan VFD is obtained by the following expression:

$$\Delta c_2(k) = -\Delta c_1(k) \cdot \frac{\Delta G_{\max}}{\Delta L_{\max}} \tag{5}$$

where $\Delta G_{\max}$ and $\Delta L_{\max}$ are the water flow and the air flow ranges, respectively.

Constraints must also be considered to establish the limits of the air and water flows. The limits of the water flow were established according to the design specifications of the equipment (Table 1) and according to the data used during the parameter estimation phase (L_{min} = 1560 m³/h and L_{max} = 2600 m³/h). The maximum value adopted for the air flow is based on the design conditions (G_{max} = 34248 m³/min and 25% of this value (G_{min} = 8562 m³/min) was assumed for the minimum air flow).

3.1. Closed-loop analysis

The plant was simulated starting from its steady-state and a sequence of disturbances in the process heat load and in the wet bulb temperature was applied. The disturbance sequence of the wet bulb temperature was defined based on data from a typical day of operation. The period was 0:00 AM to 6:00 PM with sampling time of two hours. The heat load disturbance sequence was defined through an energy balance using historical data of inlet and outlet temperatures and water flow. Thus, the wet bulb temperature in the period considered ranged from 23.5 to 25.5 °C and the heat load from 65 to 95 GJ/h. The effect of these disturbances on the efficiency, effectiveness and exit water temperature were used as a base case for comparison with the closed-loop results. The closed-loop strategy comprised an override/split-range combination in order to maximize the cooling tower performance through the use of two PID controllers considering simultaneously the control of efficiency and effectiveness, variables with opposite behaviours as shown earlier. The diagram of this override/split-range strategy is presented in Figure 1.
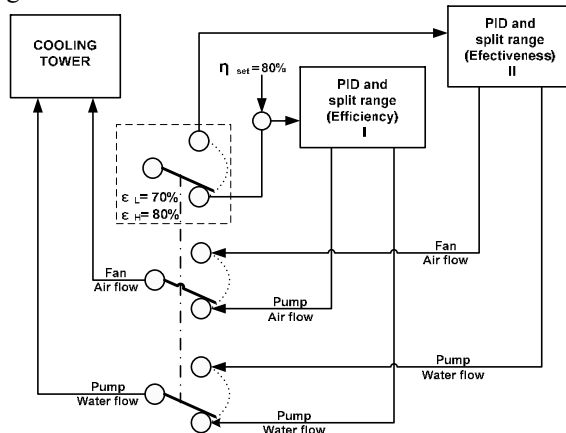
Figure 1 - Override/split-range strategy designed to control efficiency and effectiveness.

Here, whenever effectiveness drops below its minimum threshold value (low effectiveness, ε_L=70%), the control is switched to the PID controller associated to the maximization of the effectiveness. On the other hand, if the effectiveness value reaches its maximum threshold value (high effectiveness, ε_H=80%), the PID controller associated to the efficiency control is activated. Despite the fact that the air and water flows are manipulated simultaneously through a split-range strategy, the effect of the water flow is more expressive in the efficiency control and the same is valid for the air flow-effectiveness interaction. Figures 2a and 2b present the behavior of manipulated and controlled variables. Considering the infeasibility of keeping efficiency and effectiveness simultaneously at their set-points, the override/split-range strategy proposed is capable of monitoring both outputs and to assure the values are maintained within satisfactory levels.
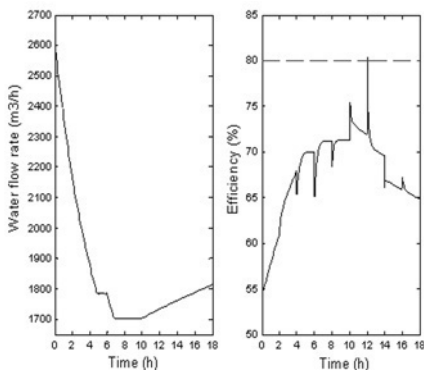
Figure 2a - Water flow and efficiency.
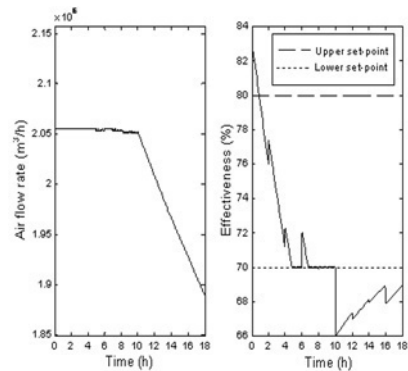
Figure 2b - Air flow and effectiveness.

The PID tuning parameters were selected through an optimal tuning method which involved the solution of an optimization problem to minimize the deviations between the controlled variable and its set-point (Alfano Neto and Embiruçu, 2000). In order to assure a workable profile for the manipulated variables based on the limits imposed by the physical components of pumps and fans, constraints were considered in the optimization problem, assuring that both water and air flow did not surpass them. A feasible performance for the manipulated variables can be verified and the control is able to cope with the disturbances to keep the controlled output close to its set-point. Moreover, it can be seen that the closed loop test presented in this work comprise a mixed control problem (regulatory and servo). The initial conditions of the controlled variables are not equal to the set-point established in each case. The potential to water and energy saving is summarized on Table 3.

Table 3 – Water and energy consumptions over 18 hours of operation

	Override/split-range strategy	Open-loop results
Water Consumption (m^3)	919.3	921.6
Energy consumption (pump) (kWh)	4332.0	11297.0
Energy consumption (fans) (kWh)	3469.8	3668.6
Total energy consumption (kWh)	9128.2	14966.0
Monetary savings (US$)	560	-

4. Conclusions

Open and closed-loop analysis of a counter flow wet cooling tower coupled with a shell and tube heat exchanger which represents the process heat load is presented. Apart from model parameter estimation using data from a commercial tower, the open loop analysis showed some particularities of the cooling tower efficiency, as first order behaviour, inverse response, nonlinearities, over-damped behaviour and opposing behaviours for the manipulated variables. The closed-loop analysis tested override/split-range strategy designed to control efficiency and effectiveness. The results demonstrate the good performance of the override/split-range strategy control in regulatory problems. Thus, potentials energy saving was estimated in US$ 250000/year. The results presented here show that significant savings can be achieved through appropriate control of certain process variables associated to the operation of a cooling tower without needing considerable investment.

References

A.N. Nahavandi, J.J. Oellinger, 1977, An improved model for the analysis of evaporative counterflow cooling towers, Nuclear Engineering and Design, v.40 (2), p.327-336.

J.W. Sutherland, 1983, Analysis of mechanical-draught counterflow air/water cooling towers, J Heat Transfer, v.105 , p.576–583.

J.R. Khan, S.M. Zubair, 2001, An improved design and rating analyses of counter flow wet cooling towers, Int. J. Heat Mass Transf, v.123, p.770-778.

A.K. Majumdar, A.K. Singhal, D.B. Spalding, 1983, Numerical modeling of wet cooling towers – Part I: Mathematical and physical models, Journal of Heat Transfer, v.105(4), p.736-743.

A.A. Dreyer, P.J. Erens, 1996, Modeling of cooling Tower splash pack, Int. J. Heat Mass Transf, v.39 (1), p.109-123.

H.T.A. El-Dessouky, A. Al-Haddad, F. Al-Juwayhel, 1997, A modified analysis of counter flow cooling towers, ASME J Heat Transf, v.119 (3), p.617–626

M.S. Söylemez, 2004, On the optimum performance of forced draft counter flow cooling towers, Energ Convers Manage,v. 45, p.2335–2341.

M. Hosoz, H.M. Ertunc, H. Bulgurcu, 2007, Performance prediction of a cooling tower using artificial neural network, Energy Conversion and Management,v. 48 (4), p.1349-1359.

M.A. Younis, M.A. Fahim, N. Wakao, 1987, Heat input-response in cooling tower-zeroth moments of temperature variations. J. Chem. Eng. Japan, v.20 (6), p.615-618.

W.F. Ramirez, 1997, Computational Methods for Process Simulation, Butterworths, 2nd edition.

S.P Fisenko, A.I. Petruchik, 2004, Towards to the control system of mechanical draft cooling tower of film type, Int. J. Heat and Mass Transfer, v.48, p.31-35.

J.C Van Dijk, C.G. Van Poll, P. Poldervaart, 1985, Investment in cooling tower control pays big dividends, Process Engineering, v.66 (10), p.57-60.

K.A. Fonstad, 1988, Energy saving through control of cooling tower air flow, Energy Engineering, v.85 (3), p.4-16.

M.-L. Chen, H.-Y. Liang, 2005, Application of Fuzzy Control in Cooling System's Save Energy Design, Int.Journal of Power and Energy Systems, v.465, p.340-344.

M.R. Stout Jr, J.W. Leach, 2003, Cooling tower fan control for energy efficiency, Energy Engineering, v.99 (1), p.7-31.

C. Alfano Neto, M. Embiruçu, 2000, Tuning of PID Controllers: an Optimization-Based Method, In Digital Control 2000: Past, Present and Future of PID Control, p.415-420.

1455

Industrial Experience in the Deployment of Real Time Online Energy Management Systems

Diego Ruiz [a], Carlos A. Ruiz [b]

Soteica Europe, S.L.; Av. Parc nº 2; E-08940 Cornellà (Barcelona), Spain;
Tel: +34 93 375 3503; http://www.soteica.eu
[a] *diego.ruiz@soteica.com*
[b] *carlos.ruiz@soteica.com*

Abstract

This paper presents real industrial examples in which the whole utilities system of a production Site (i.e., steam, fuels, boiler feed water and electricity) is optimized with a real time online, industrially well established software.

Experiences gained during more than 20 years of industrial projects deployed worldwide are commented. Main project steps are explained and critical details to be taken into account to assure successful use and proper technology transfer are presented. The optimization objective is the overall utilities system cost reduction and takes into account the constraints associated with the existing equipment, fuels and electricity pricing and contracts, including emissions limits, quotas and rights. The energy management system models are executed and optimized at a scheduled frequency, fed with online, real time data, flowing into and out the program using the standard OPC protocol.

Besides the optimization, Key Performance Indicators (KPIs) are also calculated and sent back to the Site Plant Information System or DCSs for Operations and Management use.

Application examples and results corresponding to projects implemented worldwide in refineries and chemical plants are presented and commented.

Keywords: energy costs reduction, real optimization, energy management

1. Energy Management Systems

Modern industrial facilities operate complex and inter-related power and utilities systems. Tighter and increasingly restrictive regulations related to emissions are also imposing constraints and adding complexity to their management. Deregulated electric and fuels markets with varying contracted prices (seasonal or daily) also introduce additional challenges.

Production Departments usually have the responsibility for the operation of the facility power system but, although Operators are instructed to minimize energy usage and usually tend to do it, the power system is seen as a subsidiary provider of the utilities needed to accomplish the production target, whatever it takes to generate it.

Industrial facilities like Refineries and Petrochemicals are becoming increasingly aware that power systems need to be optimally managed because any energy reduction that Operations accomplish in the producing Units could eventually be wasted if the overall power and utilities system cost is not properly managed.

In parallel, process engineers have always attempted to develop some kind of tool, many times spreadsheet based, to improve the way utilities systems are operated. The

evolution from plant information scattered throughout many islands of automation to unified and centralized Plant Information Systems was a clear breakthrough for such work. The long term, facility wide Plant Information System based historians constitute what is known as an *enabling technology*, because they became the cornerstone on top of which many other applications are built.

Process engineers used the internal optimizers or solvers provided with the spreadsheet software in an attempt to optimize their systems. One of the authors went through a very similar path when he was a young process engineer at a Petrochemical Complex. After many years of exposure to a lot of manufacturing sites worldwide, he found that almost all Process Engineering Departments had in use an internal spreadsheet with which several process engineers worked when on duty at the power house or utilities unit. The spreadsheets wildly evolved during the years and became extremely complex and "fragile" as hundreds or thousands of tags were added and, very often, without any documentation. Raw tag data becomes usually contaminated with unexpected problems and sometimes hard to identify and filter errors. The use of such a tools for real time optimization was seldom a real success.

1.1. Real time, Intelligent Systems

It was becoming more and more clear that a certain kind of *intelligence* should be added to those energy management tools in order to produce good results in a consistent way, dealing with real time information potential errors and maintained evergreen and usable for long periods, with minimal engineering effort.

The authors found that, for the Process Industry, the definition of an *intelligent system* is generic and not very well defined. The industry usually calls *intelligent* to any piece of software that helps to automate the decision making process, efficiently controls a complex process, is able to predict properties of products or process variables, alerts to prevent hazardous situations or, in last instance, optimize process or business economics. For the practical engineers, the definition of an intelligent system is factual, not methodological. The above mentioned systems comply, up to certain extent, with one of the classical definitions of intelligence (N. Wiener, 1948): the intelligent behavior is a consequence of certain feedback mechanisms, based on the acquisition and processing of information to accomplish with a certain objective.

A coherent engineering environment providing all the needed tools within a single shell was a real need. During the past 20 years, Visual MESA optimization software evolved from the earlier text based, offline application of the 1980's to an online, real time, graphical user interfaced, highly sophisticated *intelligent* system. It is today considered as the industry standard Energy Management System (EMS) real time online optimizer (D. Nelson et al., 2000). It has been widely implemented in the processing industry and it is applied routinely to reduce the cost of operating the energy systems at power, chemical, petrochemical, and refinery plants worldwide.

1.2. Energy Management System Description

A detailed model of the energy system (fuels, steam, electricity, boiler feed water and condensates network) is built within the EMS graphical user interface and it is continuously fed with validated, real time data. It includes all the actual constraints of the site and decision variables for their operation. Optimization is configured to minimize the total energy cost. The model writes back its results to the Real Time Data Base (Plant Information System). It also provides reliable data that helps to audit the energy productions and usages within the site energy system, and in that way wastes can be detected and eliminated. Greenhouse emissions are also taken into account.

1.3. Real Time, Online Systems

1.3.1. Manual, Open Loop versus Closed Loop Operation

Although the Operators still need to close the loop manually, EMS projects have proven to save substantial amounts of money with very fast pay-back. Even more economic benefits can be obtained if some of the manual optimization handles are automated under a closed loop scheme (D. Uztürk et al., 2006).

There are several and important additional advantages of using EMS's for online, real time, closed loop optimization because it increases the benefits already obtained in open loop, especially when fuels and power prices are market driven and highly variable. Several implementations of this kind have already been performed (Wellons et al., 1994 and D. Uztürk et al., 2006).

1.3.2. Online capabilities

The online capabilities are a relevant portion of the software structure and key to a successful closed loop implementation. A proper software tool should provide standard features right out of the box. Therefore, it should not require any special task or project activity to enable the software to easily interact and cope with real time online data. The EMS based models are created from scratch acquiring and relying on real time online data. A standard OPC based (OLE for process control) protocol interface has been provided to perform a smooth and easy communication with the appropriate data sources, such as a distributed control system (DCS), a plant information system, a historian or a real time database.
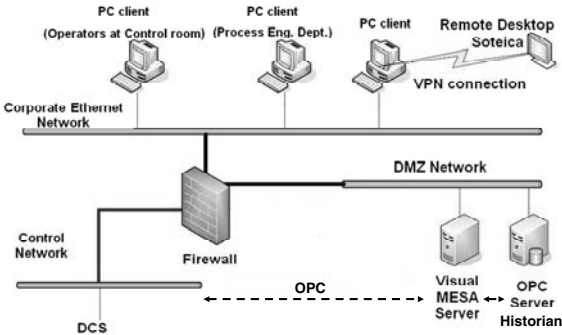
Figure 1: Installation Architecture for Closed Loop Implementation

Fig. 1 shows the typical installation architecture for closed loop real time optimization, including the proper network security layers and devices, for example firewalls and demilitarized zones (DMZ) domains.

1.4. Optimization Variables and Constraints Configuration for Closed Loop Optimization

Building a model that realistically represents the utilities and energy system topology, needs to include all the optimization variables and constraints and, at the same time, include all the system economic details, especially the fuels and electricity contractual complexity.

Optimization variables are those where some freedom exists regarding what value they might take. For example, the steam production rate at which a particular boiler operates is a free choice as long as the total steam production is satisfied, thus the most efficient boiler's production can be maximized. They can be continuous or discrete.

Constrained variables are those variables that cannot be freely chosen by the optimizer but must be limited for practical operation. Some of them are direct equipment

constraints, but a few can be defined as abstract constraints, where the variables are not directly measured in the system or they are not a related of a particular of equipment.

2. Project Activities

An Energy Management System (EMS) Implementation project is executed in 9 to 12 months. The main project steps are discussed below.

2.1. Required Information

After the Purchase Order is issued, a document would be submitted to the Site with all the informational requirements for the EMS project sent it to the project owner. By project owner we understand a Site engineer who, acting as a single interface, will provide the needed information and coordinate all the project steps.

2.2. Kick-Off Meeting

Prior to the Kick-Off Meeting, the provided information will be reviewed to have a better understanding of the Site facilities and process. Additional questions or clarifications would be sent to the Site regarding particular issues, as required. During the week of the on-site Kick-Off Meeting, all information would be reviewed with the Site staff, and additional information required for building the model would be requested, as needed. At that time, the optimization strategy would also be discussed.

2.3. EMS Software Installation

The software is then configured and licensed on the EMS server PC. It would also be connected to the OPC server. Remote access to the model would also need to be made available at this time and would need to be available throughout the rest of the project.

2.4. Functional Design Specification

With the information provided during the Kick-Off meeting, a Functional Design Specification document would be prepared, revised by both parties in concert, and then approved by the Site.

2.5. Visual MESA Model Building and Optimization Configuration

During this stage, the model and the report are built working remotely on the EMS server. The model grows with access to online real time data. The second trip to the facility would occur during this stage and would be used for mid-term review of the model and optimization. Upon model approval, a month-long testing period would commence. The EMS would run routinely, but optimization recommendations would still not be implemented by the operations staff. A base line could be obtained based on the cost reduction predicted by the optimizer during this period, in order to compare with the full implementation of the suggestions at the end of the project.

2.6. Optimization Startup

Site engineers would then train the operations staff to use Visual MESA and to implement the recommendations. Continuing in this period, operations staff would begin implementation of the optimization recommendations. Project developing staff would return to the Site facility a third time to review implementation of the optimization recommendations and make any final adjustments to the model, as required. Project documentation would be provided and a benefits report would be submitted.

3. Key Performance Indicators (KPI's)

In addition to the real time online optimization, during the EMS project appropriate energy performance metrics can also be identified and performance targets can be set.

Those metrics are usually known as Key Performance Indicators (KPI's) and can be related to:

- High level KPI's that monitor site performance and geared toward use by site and corporate management. For example: Total cost or the utilities system, predicted benefits, main steam headers imbalances, emissions, etc.
- Unit level KPI's that monitor individual unit performance and are geared toward use by unit management and technical specialists. For example: plant or area costs, boilers and heaters efficiencies, etc.
- Energy Influencing Variables (EIV's) that are geared towards use by operators. For example: Equipment specific operation parameters, like reflux rate, transfer line temperatures, cooling water temperature, etc.

They are calculated by the EMS software and written back to the Plant Information System.

4. Project Examples

The authors implemented several dozens of projects, all of them currently in service around the world. Some of them are described in D. Ruiz et al., 2005, 2006 and 2007, S. Benedicto et al., 2007, J. M. García Casas et al., 2007 and M. Kihn, et al 2008.

Two examples will be commented; the first one is an open loop implementation. The second belongs to a closed loop system.
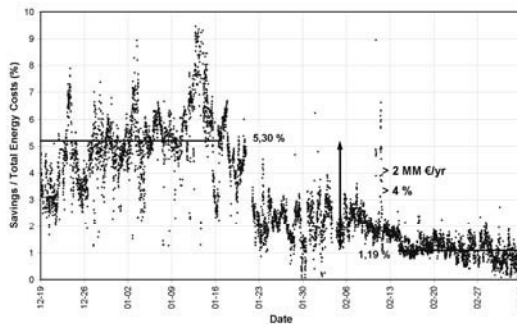
Figure 3. Energy cost reduction evolution by using an online energy management tool

4.1. *First Example*

The first example corresponds to the energy system of a Spanish refinery plus an olefins unit (D. Ruiz et al., 2006). In order to accurately evaluate the economic benefits obtained with the use of this tool, the following real time test has been done:

- *First month:* Base line, The EMS being executed online, predicting the potential benefits but no optimization actions are taken.
- *Second month:* Operators trained and optimization suggestions are gradually implemented.
- *Third month:* Optimization recommendations are followed on a daily basis.

Fig. 3 shows the results of this test. Over that period, in 2003, 4% of the energy bill of the Site was reduced, with estimated savings of more than 2 million €/year.

4.2. *Second Example*

The second example corresponds to a Dutch refinery where the EMS online optimization runs in closed loop, the so-called energy real time optimizer (D. Uztürk et al., 2006).

Typical optimization handles include letdowns, load boilers steam flow, gas turbine generators/steam turbine generators power, natural gas intake, gas turbine heat recovery, steam generators duct firing, extraction of dual outlet turbines, deaerator pressure, motor/turbine switches, etc.

Included constraints are the steam balances at each pressure level, boiler firing capacities, fuel network constraints, refinery emissions (SO_2, NO_x, etc.) and contract constraints (for both fuel and electric power sell/purchase contracts).

Benefits are reported to come from the continuous load allocation optimization between boilers, optimised extraction/condensing ratio of the dual outlet turbines, optimised mix of discretionary fuel sales/purchase, optimised gas turbine power as a function of fuel and electricity purchase contract complexities (trade off between fuel contract verses electricity contract penalties).

5. Future Perspectives

Although wide opportunities still exist for a growing number of real time online Energy Management Systems executed in open loop, an increased number of Closed Loop applications are expected in the near future. This evolution will bring additional economic benefits to the existing user base, especially when fuels and power prices are market-driven and highly variable. High frequency optimization opportunities that cannot be practically addressed by manual operating procedures would be captured and materialized. Of critical importance is having a robust and mature solver that reliably converges in a reasonable period of time in order to ensure buy-in from Operations for the continuous use of the system.

More focus on key process side operations, when tightly related with the Energy Network, will be also necessary. In addition to the Refining and Petrochemical industries, who were the early adopters of this kind of technology, other industries will take advantage of the real time energy management systems as well.

References

N. Wiener, 1948, Cybernetics, or Control and Communication in the Animal and the Machine, 2nd ed. of 1961, Cambridge, Mass.: MIT Press

M.C. Wellons, A.V. Sapre, A. I. Chang, T. L. Laird, 1994, On-line Power Plant Optimization Improves Texas Refiner's Bottom Line, Oil & Gas Journal, May 16

D. Nelson, G. Roseme, S. Delk, 2000, Using Visual MESA to Optimize Refinery Steam Systems, AIChE Spring Meeting, Session T9013, Georgia, USA

D. Ruiz, C. Ruiz, J. Mamprin, Depto. de Energías y Efluentes Petronor, 2005, Auditing and control of energy costs in a large refinery by using an on line tool, European Refining Technology Conference (ERTC) Asset Maximisation, Budapest, Hungary

D. Ruiz, C. Ruiz, D. Nelson, G. Roseme, M. Lázaro, M. Sartaguda, 2006, Reducing refinery energy costs, Petroleum Technology Quarterly (PTQ), Q1 2006, Pages 103-105

S. Benedicto, B. Garrote, D. Ruiz, J. Mamprin, C. Ruiz, 2007, Online energy management, Petroleum Technology Quarterly (PTQ), Q1 2007, Pages 131-138

J. M. García Casas, M. Kihn, D. Ruiz, C. Ruiz, 2007, The Use of an On-line model for Energy Site-Wide Costs Minimisation, , European Refining Technology Conference (ERTC) Asset Maximisation Conference, Rome, Italy

D. Uztürk, H. D. Franklin, J. M. Righi, A. T. Georgiou, 2006, Energy System Real Time Optimization, NPRA Plant Automation and Decision Support Conference, Phoenix, USA

D. Ruiz, C. Ruiz, D. Nelson, 2007, Online Energy Management, Hydrocarbon Engineering, September, Pages 60-68.

M. Kihn, D. Ruiz, C. Ruiz, A. García Nogales, 2008, Online Energy Costs Optimizer at Petrochemical Plant, Hydrocarbon Engineering, May, Pages 119-123

Optimizing Control Action Online Using a Neural Model and the Solver of an Electronic Worksheet

Tatiana L. Fujiki[a], Márcia R. G. Sanzovo, Manuela S. Leite, Flavio V. Silva,
Ana M. F. Fileti[a]

[a] *Department of Chemical Systems Engineering, School of Chemical Engineering -
University of Campinas - Av. Albert Einstein, n. 500, P. O. Box 6066, 13083-970,
Campinas - SP, Brazil.*

Abstract

The worldwide market requirements for high standard products and safe and
environmentally friendly processes enforce chemical industries to look for controllers
that can afford nonlinearities and transient behavior of chemical processes. Intelligent
systems based on Artificial Neural Network (ANN) have been studied as alternative
solutions for ill-defined plants or nonlinear and transient systems. In this work, the
algebraic equations of a neural model were implemented in a Microsoft Excel
worksheet. As a case-study, this alternative controller was successfully implemented to
maintain the bulk temperature of a bromelain precipitation tank at 5°C.

Keywords: artificial neural network, Fieldbus interface, digital process control,
Microsoft Excel Solver.

1. Introduction

A 2-year industrial investigation into the application of ANNs in process control was
summarized in 2001 by Lennox et al. Among their studies, the required accuracy in a
polymer quality control system was successfully achieved, by integrating an ANN
within a model based control algorithm. The controller was implemented online by
interfacing an IBM personal computer with a 286 processor to the existing data
acquisition and control system. Gadkar et al. (2005) proposed the application of neural
networks for monitoring the state variables of an alcoholic fermentation tank. Through a
weight updating algorithm, the neural model was implemented online and the
comparative results showed that the error generated by the model with weights
adaptation was smaller than that obtained by the model without weights adaptation.
Effectiveness of neural networks was also investigated by Fileti et al. in 2006, during
the online implementation of an inverse neural model to automatically control a basic
oxygen steel-making (BOS) process. An increase in productivity was observed when
applying the neural model to predict end-point temperature and carbon percentage in the
steel at the end of the batch period. Gonzaga et al. (2009) used a feed forward ANN to
develop a soft-sensor that estimated online the polyethylene terephthalate (PET)
viscosity, for application in a control system. The integration between the developed
soft-sensor and the supervisory system SETCIM (AspenTech) was performed using
scientific programming languages, such as FORTRAN and C. As can be seen, ANNs
have been largely employed for nonlinear process modeling because of their ease of
construction and implementation (since the need for mechanistic modeling from first
principles is suppressed). They have been proved effective in modeling and predicting

process variables. The use of a Microsoft Excel worksheet for online implementation instead of other commercial software or computational programming is proposed in this work due to its friendly interface and applicability.

2. Materials and Methods

2.1. Neural Model

Using the Neural Network Toolbox of MATLAB 7.0, a dynamic neural model was developed so that its optimized parameters could be determined. The input layer consisted of the current values (k) of the measured variables of the process. One sampling time was considered between the instants k and k+1 used in the model. In the hidden layer, an activation function was applied to twice the number of variables of the input layer. Another activation function was applied to the output layer, which predicts the one-step-ahead controlled variable, PV_{k+1}. The Levenberg-Marquardt algorithm was employed to train the multilayered feed forward network and the training procedure was performed using open-loop runs. The open-loop data set was obtained by gathering a wide range of values of the input layer variables, including the whole network action domain. The process dynamics was observed through an open-loop run with the manipulated variable at a fixed point. From there, the step disturbances in the manipulated variable, *MV*, were planned so that the controlled variable behavior could be monitored from several runs and this database was employed to train the neural network. The database was split in two sets: 75% and 25% for training and tests, respectively. Furthermore, closed-loop runs with a fuzzy controller were also used for the tests, since the neural model was expected to have good response in closed-loop, with the implementation of the optimizer for the process control. The neural model performance was assessed through dispersion plots of the testing runs, with a desirable result being represented by a slope coefficient of the linear fitting of the dispersion plots (network output *versus* target vector) close to the unity and an interception coefficient around zero. Following the offline tests, online validation was performed, in order to verify the prediction capacity of the model to be implemented.

2.2. Using an Electronic Worksheet as a Controller

The optimized weights and biases were inserted in an electronic worksheet (Microsoft Excel) to reproduce the algebraic equations of a neural model. The input vector with the process data was normalized, according to Equation 2.1, using the maximum and minimum values of the variables:

$$pn = 2\frac{(p - \min p)}{(\max p - \min p)} - 1 \qquad\qquad Equation\ 2.1$$

The normalized signal from the input layer was multiplied by the weights between input and hidden neurons and each result was added to the correspondent bias. This number, called *n*, was then processed in the hidden layer according to the hyperbolic tangent activation function (Equation 2.2). Similarly, the output layer processed the signal from the hidden layer, *a*. In the output neuron a simple linear activation function was used.

$$a = \tan sig(n) = \frac{2}{(1 + \exp(-2n))} - 1 \qquad \text{Equation 2.2}$$

The computed output value was then denormalized, by isolating p in Equation 2.1, in order to be compared to its set-point. The quadratic error of the one-step-ahead controlled variable (PV_{k+1}) relative to its set-point was defined as the objective function to be minimized by the solver (Equation 2.3). The solution to be implemented in the plant was found by means of the *quasi*-Newton method of generalized reduced gradient, available in Excel, by changing the manipulated variable value, *MV*.

$$\min_{MV}\left(PV_{(k+1)} - PV_{sp}\right)^2 \qquad \text{Equation 2.3}$$

Subject to the constraints of Equations 2.4, 2.5 and 2.6:

$$MV \geq MV_{min} \qquad \text{Equation 2.4}$$

$$MV \leq MV_{max} \qquad \text{Equation 2.5}$$

$$\left|MV_{solver} - MV_{k-1}\right| \leq \text{maximum step allowed} \qquad \text{Equation 2.6}$$

2.3. Case-Study

Bromelain is an enzyme found in pineapple, even at its mature stages. It can be recovered through precipitation with ethanol. However, this solvent may cause protein denaturation, thus the precipitation should be carried out at low temperatures, reducing molecule flexibility and hence, solvent penetration (SCOPES, 1994). The experiments were carried out in a stirred fed-batch jacketed tank. The instrumentation diagram and the pilot-plant are presented in Figure 2.1.
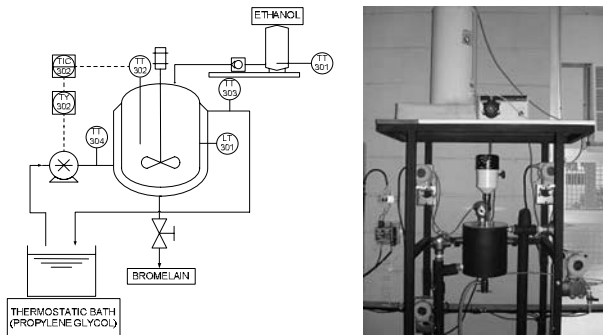
Figure 2.1: Instrumentation diagram and prototype of the bromelain precipitation process.

Intelligent instruments such as temperature transmitters were used for monitoring the process variables. The data acquisition and process control were performed by means of a digital Fieldbus network architecture (bus with spurs topology), configured by Syscon Software 7.0 (Smar). For the experimental tests, 150 mL samples of pineapple juice

were fed into the tank and the ethanol (99.5%) at room temperature was continuously fed by a micropump, until the liquid volume reached 750 mL. The coolant (propylene glycol) flow rate was manipulated through a variable speed pump. Syscon and Excel communicated with the supervisory software (Indusoft Web Studio) through Ole for Process Control (OPC) protocol.

3. Results and Discussion

3.1. Neural Model

The operating variables for the input layer, measured every 4 seconds (sample time), were chosen as follows: ethanol temperature ($T_{alc,k}$) has great influence in the precipitation temperature – the higher the ethanol temperature, the higher the overshoot obtained; coolant inlet ($T_{in,k}$) and outlet ($T_{out,k}$) temperatures provide information on the heat exchange in the tank jacket; since coolant pump speed (MV_k) is the manipulated variable, its value will determine the controlled variable response; the liquid level (L_k), represented by the liquid volume, provides the ANN with information on the run time, thus distinguishing equal input vectors that correspond to different output vectors; pump speed variation (ΔMV_k) indicates to the ANN which step disturbance in pump speed caused the given output; since bulk temperature ($T_{bulk,k}$) variation determines the manipulated variable action, it works as a reference for the ANN prediction of the one-step-ahead bulk temperature ($T_{bulk,k+1}$), that was chosen as the output variable of the ANN. In the hidden layer a hyperbolic tangent activation function was applied to 14 nodes. The structure of the neural model developed is shown in Figure 3.1.
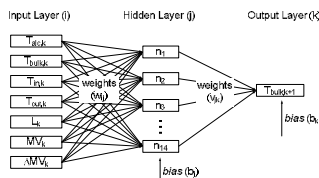
Figure 3.1: Topology of the neural model

In Figure 3.2 the results for the open-loop and the closed-loop offline tests of the neural model in MATLAB are presented.
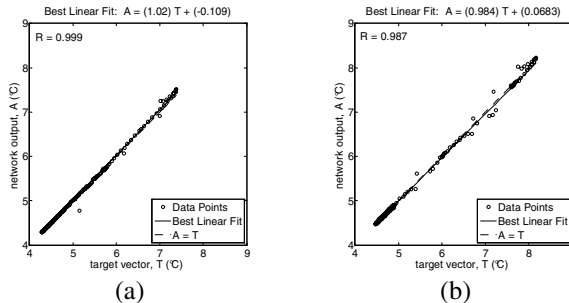
Figure 3.2: Dispersion plot of the network output and the target vector for an open-loop run (a) and for a closed-loop run (b), for the test of the neural model.

Offline tests with unseen data proved that the ANN successfully predicted the tank temperature, as can be seen by the agreement between real and predicted values of temperature. Both linear fits presented slope coefficients close to the unity and interception coefficients around zero, approaching to the ideal curve.

3.2. Online Validation and Closed-Loop Runs

At this stage, the bulk temperature was monitored during a closed-loop run with the PID controller ($K_c = 35\%/°C$, $\tau_I = 28s$ and $\tau_D = 7s$). Simultaneously, the neural model predicted the one-step-ahead bulk temperature. At the end of the experiment both temperatures were compared and the result is presented in Figure 3.2 (validation curves). The results showed that the ANN was capable of learning the inherent nonlinearities and also successfully predicted the bulk temperature in online mode, thus being considered suitable for the application with the solver. The neural model coupled with the Microsoft Excel Solver was used as a temperature controller and the experimental results are shown in Figure 3.2 (alternative controller curve). The manipulated variable, to be changed by the solver was subject to the following constraints: a range of 0 to 100% of speed variation was set, in function of its own actual limitation (Equations 2.4 and 2.5). Also, the manipulated variable action was smoothed by restricting the value of its step (the difference between the solver solution and the present control action) to 35%, based on the PID controller gain. Preliminary tests showed that these constraints were not sufficient to keep the bulk temperature around set-point: cooling outperformed heating after 200 s. To prevent the controlled variable from leaving the training range, an additional constraint was added: the pump was turned off when the bulk temperature would reach 4.9ºC.
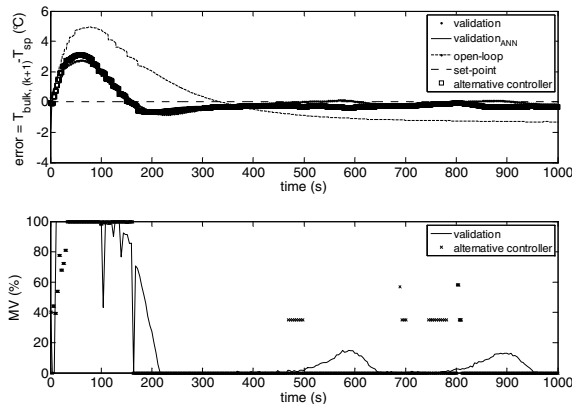
Figure 3.2: Results from the online validation and from the closed-loop run, in comparison with the open-loop run.

In Figure 3.2, it is shown that the developed alternative controller was able to maintain the controlled variable around set-point ($5°C$), with good temperature prediction. The overshoot observed in the first 200 seconds is due to the low liquid level in the tank,

which caused the heat exchange area to be at a minimum. It can also be explained by the dissolution heat produced during the ethanol addition.

In Table 3.1, a numerical comparison is presented:

Table 3.1: Comparison between PID and alternative controllers and open-loop run.

Controller	Overshoot (°C)	Rise time (s)	ITAE	Energy consumption (kJ)
PID	2.76	164	157843.7	143
Alternative	3.10	156	166716.0	121
Open-loop	4.93	334	597343.5	215

It can be noted that all the analyzed parameters for both PID and alternative controllers were smaller than the parameters obtained for the open-loop run (pump speed fixed at 40%), showing the need for process control. The low overshoot is important to avoid protein denaturation by the solvent. Smaller rise time values allow the process to be controlled faster and the lower energy consumption translates into financial savings.

4. Conclusions

This work regarded the development of an alternative controller for a semi-batch process, using artificial neural networks modeling coupled with a solver of an electronic worksheet. The neural model predicted the controlled variable with good accuracy and it was successfully used along with the solver to maintain the bulk temperature around set-point, with low energy consumption. The developed controller gathered the benefits of the artificial neural networks in affording nonlinearities and the accessibility of an electronic worksheet, making this methodology a promising new way to face complex process control problems, without spending efforts unnecessarily in mathematical modeling. Broadening the range of applicability of this method to a wide range of industrial processes is a very relevant topic of research, since industries of different fields can apply this solution to many of their problems.

References

A. M. F. Fileti, T. A. Pacianotto, A. P. Cunha, 2006, Neural modeling helps the BOS process to achieve aimed end-point conditions in liquid steel, Engineering Applications of Artificial Intelligence, v. 19, p. 9-17.

K. G. Gadkar, S. Mehra, J. Gomes, January 2005, On-line adaptation of neural networks for bioprocess control, Computers and Chemical Engineering, v.29, p. 1047-1057.

J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, R. Maciel Filho, 2009, ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process, Computers and Chemical Engineering, v. 33, p. 43-49.

B. Lennox, G. A. Montague, A. M. Frith, C. Gent, V. Bevan, 2001, Industrial application of neural networks – an investigation, Journal of Process Control, v. 11, p. 497-507.

R. K. Scopes, 1994, Protein Purification: Principles and Practical, 2.ed, New York: Springer Verlag, 329 p.

Acknowledgments

Ash Deposits Monitoring in a Convective Heat Transfer Section of a Kraft Recovery Boiler

Gustavo M. de Almeida[a], Song W. Park[b]

[a]*Federal University of Sao Joao del-Rei*
Rod. MG 443, Km 07, Fazenda do Cadete, 36.420-000, Ouro Branco, MG, Brazil
[b]*Department of Chemical Engineering, Polytechnic School, University of Sao Paulo*
Av. Prof. Luciano Gualberto, 380/Tv. 3, Butanta, 05.508-900, Sao Paulo, SP, Brazil

Abstract

A reliable monitoring of abnormal situations is still a challenge in industrial chemical processes. The earlier its detection and diagnosis, the greater the possibility of mitigating losses. An alternative approach to carry out these tasks is to make use of signal processing tools. In this direction, the hidden Markov model (HMM) method is used to monitor the accumulation of ash deposits along a convective heat transfer section of a boiler belonging to a Kraft pulping mill in Brazil. The identified model was able to provide the current state of the process working as a warning in case of a coming abnormal event.

Keywords: Chemical process monitoring, Signal processing tool, Hidden Markov model, Kraft pulping mill, Industrial data analysis

1. Introduction

Process monitoring actions in chemical industries are to guarantee operating, economic, safety and/or environmental aspects. The main idea, after detecting and diagnosing an abnormal situation, is to take actions to recover the process to a normal operating condition mitigating this way the effect of potential losses [1-2]. Since abnormal situations are in general incipient, deviations from normal conditions are smaller at the beginning, and hence reaching early detection without the support of a computer-based system is practically unfeasible. Most of the applications are normally based on residue metrics, and an alternative approach for accomplishing monitoring tasks is to make use of signal processing tools [3]. This work applies the hidden Markov model (HMM) technique to monitor the operation of a heat transfer section of a boiler. Therefore, this data-driven technique, which belongs to the signal processing field, constitutes an other way to develop chemical process monitoring systems. The case study is based on a chemical recovery boiler from a Kraft pulping mill in Brazil. Studies concerning the application of HMMs in the process monitoring area from computer-simulated cases are those by [4-13].

2. Hidden Markov Model (HMM)

Every chemical process is under random influences due to an inherent variability, and hence measurements of process variables may be seen as realizations of an underlying stochastic process. Then, normal operating conditions can be described by particular probability distributions, which fail in case of changes in process conditions [14]. This is the motivation for applying the hidden Markov model (HMM), a signal processing

tool, to the chemical process monitoring area, once it is capable of identifying changes of statistical nature in signals (composed by measurements of process variables in this case) over time. Such nature is useful in order to deal with both noise and disturbances. The successful applications of HMMs are in the speech processing field, including both speech recognition and speaker verification [15].

2.1. Fault Detection Tasks with HMMs

The goal of hidden Markov models is to model sequential data. Fig. 1 shows its input-output relation, in which the input is an observation sequence of T symbols (discrete case) or real vectors of same size (continuous case) ($O = \{o_1, o_2, ..., o_t, ..., o_T\}$), and the output is a likelihood value ($-log[P(O|\lambda)]$), a measure of the capacity of the model (λ) in generating O [15]. For example, a HMM representing an operating state in particular can be used to detect changes from it over time, which is characterized by more and more lower likelihood values.

$$O=\{o_1, o_2, ..., o_t, ..., o_T\} \longrightarrow \boxed{\text{HMM } (\lambda)} \longrightarrow -log[P(O|\lambda)]$$

Fig. 1. Input-output relation for HMMs: observation sequence (O) and likelihood value ($-log[P(O|\lambda)]$), respectively.

2.2. Mathematical Formulation

Hidden Markov models are a doubly stochastic process, in which the former is responsible for the state-transitions ($P(q_t|q_{t-1})$), which obeys the Markov property, whereas the latter concerns the observation-emissions ($P(o_t|q_t)$). The difference between Markov- and hidden Markov-models refers to this second process, that is deterministic to the first. This explains the *hidden* term in HMMs once the underlying Markov chain is not directly observable. Table 1 shows the three parameters of discrete HMMs, namely π, A, and B, where M_D is the number of distinct observation symbols of the emission probability distributions (one per state), and N is the size of the discrete state space. (A compact notation is given by λ, i.e. $\lambda = (\pi, A, B)$.) For the continuous case (used in this work), the B matrix is replaced by probability density functions, whose usual representation is a finite mixture of Gaussians, where o_t is the observation vector at time t, M_C is the number of mixture components per state, and c_{jk} is the mixture component (subjected to stochastic constraints), μ_{jk} is the mean vector, and Σ_{jk} is the covariance matrix, for the kth mixture component in the state j (cf. Table 1). The parameters π and A are the same as in the discrete case [15].

Table 1. Elements of discrete and continuous HMMs.

Description	Parameter	Formulation
State-transition probability distribution	$A = \{a_{ij}\}$	$a_{ij} = P(q_{t+1} = j \mid q_t = i)$, $1 \le i, j \le N$
Observation-emission probability distribution	Discrete case (DHMM) ($B = \{b_j(k)\}$)	$b_j(k) = P(o_1 = v_k \mid q_t = j)$, $1 \le k \le M_D$, $1 \le j \le N$
	Continuous case (CHMM)	$b_j(o_t) = \sum_{k=1}^{M_C} c_{jk} N(o_t, \mu_{jk}, \Sigma_{jk})$, $1 \le j \le N$
Initial state probability distribution	$\pi = \{\pi_i\}$	$\pi_i = P(q_1 = i)$, $1 \le j \le N$

3. Case Study

This study uses a chemical recovery boiler from a Kraft pulping mill in Brazil. One of its goals is to produce steam for electric power generation and heat transfer operations. Its fuel is the residual liquor originating from a previous stage. This equipment has two regions: a furnace, where the liquor combustion and the recovery of specific inorganic compounds occur, and a heat transfer section, as in power boilers, co-responsible for the steam production. This region contains three heat exchangers in series, namely super-heater (SH), convector (Conv), and economizer (Eco), as in Fig. 2. A risk situation refers to the accumulation of ash deposits, composed for sulfur- and sodium-based salts mostly sodium sulfate (Na_2SO_4), over tubes of the heat exchangers. This is mainly caused by sodium-based salts, vapor sodium, fuel droplet residues, and unburned carbon particles, which are carried in the fuel gas line. The consequences are: reduction of the efficiency in generating steam, lower disponibility of the boiler due to the blockage of the fuel gas path, and greater maintenance costs. The established practice in mills to manage it is the use of sootblowers [16-17]. Then, this study applies the hidden Markov modeling to monitor the increase of ash deposits over tubes of heat exchangers of a heat transfer section of a boiler. The considered system also includes the eletrostatic precipitator (cf. Fig. 2), an equipment used to recover the sodium and sulfur in the fuel gas line. The monitored variable, which is strongly correlated to the increase of ash deposits, is the pressure drop (ΔP) of the gas along the system.
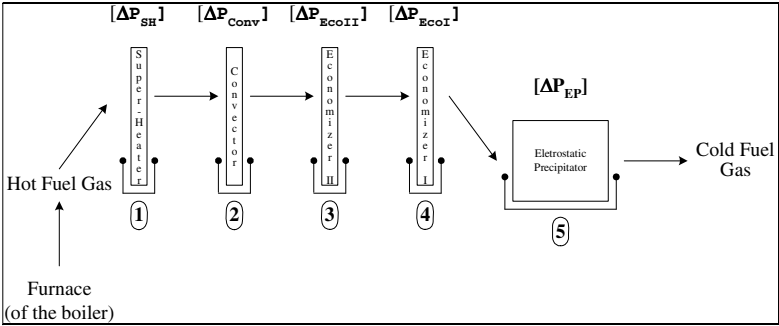
Fig. 2. Scheme for the convective heat transfer section and the eletrostatic precipitator of the boiler, with the five measurements of the pressure drop of the fuel gas.

3.1. Data Set

The data set comprehends ten months of operation, with a sampling interval of one minute. Table 2 presents the collected measurements of the pressure drops of the fuel gas along the heat transfer section until the eletrostatic precipitator (cf. Fig. 2). Other variables were collected in order to verify the operating conditions during this period.

Table 2. Measurements collected in the chemical recovery boiler.

Description	Code	Unit
Pressure drop of the fuel gas through the …		
1. Super-heater	ΔP_{SH}	Pa
2. Convector	ΔP_{Conv}	Pa
3. Economizer II	ΔP_{EcoII}	Pa
4. Economizer I	ΔP_{EcoI}	Pa
5. Eletrostatic Precipitator	ΔP_{EP}	kPa

4. Methodology

4.1. Model Identification Step

The goal is to identify a representative HMM for the boiler under a critical operating condition concerning the pressure drop of the fuel gas along the heat transfer section, which is given by ΔP_{EcoII} (the pressure drop of the fuel gas through the second economizer) $\geq$ 1000.0 Pa. A continuous HMM is used in this work (cf. Table 1) and each observation sequence is composed of 5 real vectors of size 1, i.e. $O = \{\Delta P_{SH}, \Delta P_{Conv}, \Delta P_{EcoII}, \Delta P_{EcoI}, \Delta P_{EP}\}$ (cf. Fig. 1). Every 1 minute (the sampling interval), a probability of a sequence of being generated by the reference model $(-log[P(O|\lambda)])$ is plotted in the trend plot of the likelihood function, providing information about the current state of the system. Initially, using a training subset, plenty of models are generated by varying the number of mixture components per state (M_C) from 1 up to 3, the number of states (N) between 2 and 3, and the model topology between ergodic and left-to-right. The model selection is based on the likelihood function being selected the one with the highest mean value calculated onto the validation subset. This steps makes use of the classical Baum-Welch algorithm, based on the Maximum Likelihood Estimation (MLE) principle [18].

4.2. Control Limits Construction Step

The calculus of the lower control limits (LCL) is based on the pressure drop of the fuel gas through the second economizer (ΔP_{EcoII}). This is accomplished by splitting its range of pressure drop into five intervals, namely [600 700[, [700 800[, [800 850[, [850 900[, and [900 950[, in Pa units. For each interval, with a set of representative observation sequences, the lowest likelihood value of an observation sequence in particular becomes the lower control limit to such interval. The model used to this determination is that previously identified. The classical Forward algorithm is used to calculate the model outputs [15].

4.3. Test Step

Finally, in order to verify the performance of the model, it is fed with another subset of observation sequences, simulating an on-line monitoring action. The idea is to get the behavior of the model when subjected to real operating conditions. As the boiler operates in a continually, a trend plot for the likelihood function (the model output) arises over time, which constitutes a source of information about the state of the equipment. As before, this step uses the Forward algorithm [15].

5. Results and Discussion

5.1. HMM Identification Step

The identification subset contains 2880 observation sequences being 1920 ($\approx$ 66.7%) for training and 960 for validation. The selected HMM, characteristic of an operation under a critical pressure drop of the fuel gas along the heat transfer section of the boiler, among the 12 candidate models, has 1 mixture component per state (M_C), 3 states (N), and an ergodic topology.

5.2. Control Limits Construction Step

Table 3 shows the lower control limits (LCL) for all five intervals (cf. Subsection 4.2), embracing the entire range of pressure drops of the fuel gas through the second economizer (ΔP_{EcoII}). For example, a likelihood value for an observation sequence equal to –36.1 means that it is already possible to have ΔP_{EcoII} = 850.0 Pa. It can be observed that higher the pressure drop measurement greater the likelihood value.

Table 3. Lower Control Limits (LCL).

Data Set	Lower Control Limits: A likelihood value $(-log[P(O\|\lambda)])$ equal to …	… means a possibility of already having a value for ΔP_{EcoII} equal to ...	Number of Observation Sequences
1	−3.2	950.0	377
2	−19.6	900.0	529
3	−36.1	850.0	1529
4	−51.8	800.0	1095
5	−110.3	700.0	389
6	−160.3	600.0	185

5.3. Test Step

The test subset contains 4712 observation sequences. First of all, Fig. 3(a) shows the strong correlation ($\rho = 0.985$) between the likelihood values calculated by the model onto the test subset and the measurements of pressure drop of the fuel gas through the 2nd economizer (ΔP_{EcoII}). The correlation between ΔP_{EcoII} and the measurements of pressure drop of the fuel gas through both the convector (ΔP_{Conv}) and the 1st economizer (ΔP_{EcoI}), which are the most critical heat exchangers with regard to the accumulation of ash deposits [17], is also close to 1. Hence, higher the likelihood value greater the measurement of the pressure drop of the fuel gas. (The pressure drop through the super-heater (ΔP_{SH}) is negligible.) Fig. 3(b) shows the result of brushing the likelihood values for all observation sequences for which $\Delta P_{EcoII} \geq 700.0$ Pa. It can be verified that the lowest value (–77.8) is greater than the respective lower control limit (LCL) equal to −110.3, which means that if the model output exceeds such limit it is already possible to have a pressure drop of the fuel gas through the second economizer (ΔP_{EcoII}) above 700.0 Pa. This information works as a warning about the state of the system concerning the accumulation of ash deposits. The results (not shown) are the same to all intervals for ΔP_{EcoII} (cf. Table 3), which attests that the monitoring of the accumulation of ash deposits along heat transfer sections of chemical recovery boilers using a HMM−based monitoring system is very promising. (For implementation, the sensibility concerning the correlation between likelihood and pressure drop values needs to be adjusted.)
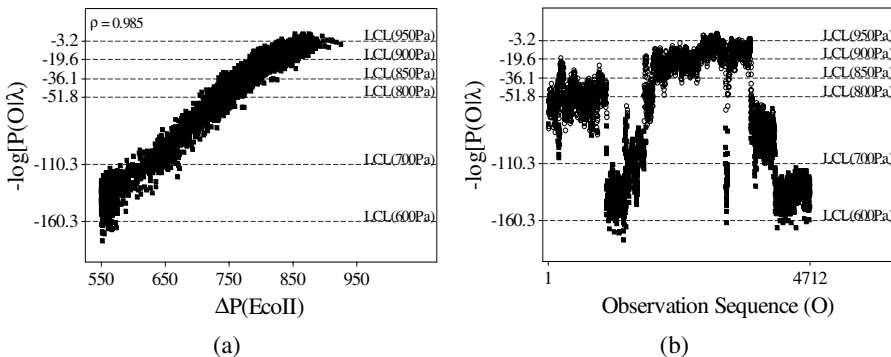
Fig. 3. (a) Linear correlation between the likelihood values calculated by the model (λ) onto the test subset ($-log[P(O\|\lambda)]$) and the measurements of pressure drop of the fuel gas through the second economizer (ΔP_{EcoII}), and (b) resulting trend plot for the likelihood function, where the values associated with observation sequences (O) for which $\Delta P_{EcoII} \geq 700.0$ Pa are highlighted by circles (○).

6. Final Considerations

A hidden Markov model (HMM), representative of high measurements of pressure drop of a fuel gas, is used to monitor the accumulation of ash deposits along a heat transfer section of a boiler. It was verified that the resulting trend plot for the likelihood function (the output of the model), which is highly correlated with the measurements of pressure drop, works as a source of information since it is able to provide the current state, and hence the tendency of the operations. In brief, HMM–based systems may work as decision support tools for monitoring actions in chemical processes.

Acknowledgments

References

[1] L.H. Chiang, E.L. Russel, R.D. Braatz, 2001, Fault detection and diagnosis in industrial systems, Springer, London.

[2] I. Nimmo, 1995, Adequately address abnormal operations, Chem. Eng. Prog., 91, 9, 36-45.

[3] R.J. Patton, J. Korbicz, S. Lesecq, 2006, Preface, Control Eng. Practice, 14, 6, 575-576.

[4] G.M. Almeida, S.W. Park, 2008, Fault detection and diagnosis in the DAMADICS benchmark actuator system – A hidden Markov model approach, Proc. of the 17th World Cong. of The Int. Fed. of Automatic Control (IFAC), Korea, 12419-12424.

[5] G.M. Almeida, S.W. Park, 2008, Process monitoring in chemical industries – A hidden Markov model approach, Proc. of the 18th European Symp. on Comp. Aided Proc. Eng., France, 25, 1-6.

[6] G.M. Almeida, S.W. Park, 2005, Fault detection in a sugar evaporation process using hidden Markov models, Proc. of the Int. Symp. on Advanced Control of Ind. Proc., Korea, 309-313.

[7] A. Bakhtazad, A. Palazoglu, J.A. Romagnoli, 2000, Detection and classification of abnormal process situations using multidimensional wavelet domain hidden Markov trees, Comp. and Chem. Eng., 24, 2-7, 769-775.

[8] J. Chen, W. Chang, 2005, Applying wavelet–based hidden Markov tree to enhancing performance of process monitoring, Chem. Eng. Sc., 60, 18, 5129-5143.

[9] W. Sun, A. Palazoglu, J.A. Romagnoli, 2003, Detecting abnormal process trends by wavelet-domain hidden Markov models, AIChE Journal, 49, 1, 140-150.

[10] F. Tokatli, A. Cinar, 2004, Fault detection and diagnosis in a food pasteurization process with hidden Markov models, The Can. J. of Chem. Eng., 82, 6, 1252-1262.

[11] J.C. Wong, K.A. McDonald, A. Palazoglu, 2001, Classification of abnormal plant operation using multiple process variable trends, J. of Proc. Control, 11, 4, 409-418.

[12] J.C. Wong, K.A. McDonald, A. Palazoglu, 1998, Classification of process trends based on fuzzified symbolic representation and hidden Markov models, J. of Proc. Control, 8, 5-6, 395-408.

[13] X. Yangsheng, G. Ming, 2004, Hidden Markov model-based process monitoring system, J. of Int. Man., 15, 3, 337-350.

[14] V. Venkatasubramanian, R. Rengaswamy, S.N. Kavuri, S.N., K. Yin, 2003, A review of process fault diagnosis – Part III: Process history based methods, Comp. and Chem. Eng., 27, 3, 327-346.

[15] L. Rabiner, 1989, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. of the IEEE, 77, 2, 257-286.

[16] T.N. Adams, W.J. Frederick, T.M. Grace, M. Hupa, K. Iisa, A.K. Jones, H. Tran, 1997, Kraft Recovery Boilers, Tappi Press, Atlanta.

[17] E.K. Vakkilainen, 2005, Kraft recovery boilers: Principles and practice, Valopaino Oy, Helsink.

[18] L.E. Baum, T. Petrie, G. Soules, N. Weiss, 1970, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat., 41, 1, 164-171.

Evaluation of Control Structures for a Debutanizer Column

Lilian R. Canabarro,[a] Mario C. M. M. Campos,[b] Enrique L. Lima[a]

[a]*Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária, CP 68502, Rio de Janeiro, RJ, Brazil 21945-970*
[b]*PETROBRAS S.A. Av. Horacio Macedo, 950, Rio de Janeiro RJ, Brazil 21941-915*

Abstract

In this work is described the choice of the control structure for a specific industrial distillation column presenting a significant discrepancy between the inventories of the top and bottom sections. The studied distillation column showed to be an excellent research field, but looking for the balance between scientific development and industrial application it is always important to take into consideration the variable "time invested". In this context, and searching for a satisfactory control structure, static analysis tools, as RGA and SVD, had been used in the characterization of a set of alternatives structures based on the primary composition control variables L, V, D and B, as well as some relations between them. With the reduction of the "time invested" kept in mind, it was used a rigorous dynamic simulator capable to deal with thermodynamic calculations for complex multicomponent mixtures, as those found in the processing units of crude oil and its fractions. Results have demonstrated the importance of rigorous dynamic simulation to solve design control problems in distillation columns which frequently presents unique characteristics and should be treated accordingly.

Keywords: Distillation control, control structure selection, dynamic simulation

1. Introduction

The interest on distillation control systems started in the decade of the 50s, significantly increasing along the time, with special emphasis in the 80s. One conclusion that can be drawn from the huge amount of technical and scientific literature produced in this period is that distillation is a difficult system to understand. The scientific research is continuously improving our knowledge on the subject, but the practice is still confronted with serious difficulties, because each distillation column seems to be different from other ones and, in such a way, requires a personalized treatment. Recently Skogestad (2007) wrote "Many books...and papers, including several of my own..., have been written on the merits of the various configurations, but it is probably safe to say that the importance of the choice of configuration (level control scheme) has been overemphasized". These words are a good example of the knowledge (scientific or general) evolution process, as they clearly express the permanent and necessary substitution of old "truths" by new ones. Certainly this subject deserves much more deep thoughts, but the idea here is only to show that in complex cases, as certainly can be the distillation columns, many of the "general truths" that had been enunciated had not been confirmed in practice, requiring more specific approaches. In his contribution Skogestad (2007) proposed a distillation column control methodology with reasonable general application, bringing great contribution for the practical side of the problem. The focus of this work is on the control structure design for a specific debutanizer

column using simple and practical tools, complemented with rigorous dynamic simulation.

The Problem

The studied system is a 26 trays naphtha stabilizer column of a crude atmospheric distillation unit. The feed, coming from the top of the atmospheric tower and introduced in tray 13th (from the top), contains 30 components, including 2% of light components (methane, ethane, propane, butanes, pentanes, etc.), water and 20 pseudo-components. The bottom product of the column is used to preheat this feed. Because it is almost totally in the liquid phase, vapor and liquid flows in the stripping section are larger than in the rectifying section, and as a consequence the thermal load to the reboiler is larger than to the condenser. There are three products in the system: combustible gas and liquefied petroleum gas (LPG) in the vapor and liquid streams that leave a partial condenser; and stabilized naphtha in the bottom stream. The control objective is to obtain a top product with the highest purity, guaranteeing the LPG quality, without reducing the quality of the stabilized naphtha. As this is produced in a large amount, significant quality variations are not expected, so the top composition control was considered a priority.

Technical-scientific literature shows that the Relative Gain Array (RGA) introduced by Bristol (1966) is still an important analysis and design tool for multivariate systems. Initially developed to determine static interaction in systems, allowing appropriate pairing between input and output variables in a control structure, its area of applicability was significantly extended later, both for static and dynamic conditions (Grossdidier et al., 1985).

Considering the stationary case and using the system gain matrix K, the RGA matrix Λ can be obtained as follows,

$$\Lambda = \left[\lambda_{ij}\right] = \left[\underline{\underline{K}}_{ij} \cdot \left[\left(\underline{\underline{K}}^{-1}\right)^{T}\right]_{ij}\right]$$

where λ_{ij} is the ratio of the static gain of the output variable i related to the input variable j, with all the other pairs in open loop, to the static gain of the same pair (i,j), with all the other pairs under perfect control. A RGA matrix close to identity indicates small interaction, determining the appropriated pairing. Another tool, sometimes complementary to the RGA, results from Singular Value Decomposition, SVD (Grossdidier et al., l985). Also considering the static case and starting from the gain matrix K, the SVD is defined as

$$K = U\Sigma V^{T}$$

where U and V are unitary matrices, and Σ is a diagonal matrix with the singular values in descending order of magnitude. The ratio of the largest to the smallest singular value defines the Condition Number, γ, a measure of the difficulty to control the associated system. As this number is dependent on variables units (scaling) it can take very different values for the same system. The correct way to take advantage of the information supplied by the condition number is to use the scaling that produces the minimum value, γ^* (scaling doesn't change system's characteristics). Research literature clearly shows limitations when evaluating a system operating in dynamic

condition using static information. However, it is important to notice that in many practical cases it is difficult to get a reasonable dynamic description of the system, limiting the use of complex elaborated theoretical methods.

An approach for calculating the gain matrices of different alternative control structures from the knowledge of only one aroused great interest in the end of the 80s and early 90s. The method introduced by Häggblom and Waller (1988) was based on transformations and relations of consistency derived from global and component mass balances. This approach seemed to be quite useful for simplifying the calculation of RGA and SDV of different control structures.

Considering only the top and bottom distillation products quality control, there are 4 primary manipulated variables available: reflux (L), distillate (D), bottom (B) and vapor (V) flowrates. From these primary variables the literature reports a large number of combinations that generates secondary variables, allowing the design of a large number of control structures. Disregarding any one of them could be a difficult task, requiring significant testing work. This is facilitated by combining the use of appropriate indices, such as RGA and γ (there are many others, with varying complexity), system knowledge and computer simulation.

2. Implementation

A set of control structures was chosen for evaluation disregarding some literature recommendations (sometimes divergent), because they are commonly based on results obtained from very simplified systems. Due to the large bottom flowrate the popular LB structure was not considered. The structures chosen were: LV, DV and (L/D, V/B). A fourth structure was also chosen where, as suggested by Skogestad (2007), a fast tray temperature loop was include in the stripping section. In this structure, identified as LV-cascade, the top composition is controlled by a cascade control system, manipulating a pump's rotor speed, where the internal loop controls the reflux. The bottom composition is controlled by another cascade control system, manipulating V, where the internal loop controls the 19th tray temperature. Table 1 shows the values of K, RGA, γ e γ^* for each control structure.

Table 1 Static gains, RGA and conditional numbers

Structure	K	Λ	γ	γ^*
LV	$\begin{bmatrix} 1.023 & -0.5806 \\ 3.88e^{-6} & -1.12e^{-3} \end{bmatrix}$	$\begin{bmatrix} 1.0018 & -0.0018 \\ -0.0018 & 1.0018 \end{bmatrix}$	1259.5	1.0892
DV	$\begin{bmatrix} -2.12 & -0.6014 \\ -6.86e^{-6} & -1.12e^{-3} \end{bmatrix}$	$\begin{bmatrix} 1.0017 & -0.0017 \\ -0.0017 & 1.0017 \end{bmatrix}$	2048.7	1.087
L/D V/B	$\begin{bmatrix} 3.70e^{-2} & -5.6202 \\ 9.45e^{-8} & -1.08e^{-2} \end{bmatrix}$	$\begin{bmatrix} 1.0013 & -0.0013 \\ -0.0013 & 1.0013 \end{bmatrix}$	79155	1.0757
LV-Cascade	$\begin{bmatrix} 2.1790 & -1.11e^{-2} \\ 8.37e^{-6} & -2.68e^{-6} \end{bmatrix}$	$\begin{bmatrix} 1.0162 & -0.0162 \\ -0.0162 & 1.0162 \end{bmatrix}$	826230	1.2887

The RGA and γ^* values indicate that the different structures should present similar behavior, in terms of insignificant interaction and good conditioning (although system scaling is weak). From this information it is possible to choose good variable pairing

and to have a feeling of the control viability of the different structures, but it is not enough to define which structure is the best. To this end dynamic simulation (gPROMS®, Multiflash®) was used, comparing the behavior of the different structures when the system is perturbed by the most common disturbances in distillation columns: changes in feed flowrate, temperature and composition. Before presenting the simulation results, it is important to shortly comment about the experience we have with the transformations method (Häggblom e Waller, 1988). Using the gain matrix of the LV structure the method was used in order to calculate the gain matrices of the other ones. As can be seen in Table 2, there was no consistence in the obtained results.

Table 2 RGA by transformations from LV

Structure	Λ
L/D V/B	$\begin{bmatrix} -2.4442 & 3.4442 \\ 3.4442 & -2.4442 \end{bmatrix}$
DV	$\begin{bmatrix} 1.0002 & -0.0002 \\ -0.0002 & 1.0002 \end{bmatrix}$

Although the result obtained for DV was similar to the one obtained from data specifically acquired for this structure (Table 1), the result obtained for (L/D,V/B) was completely different. This could be an indication that the assumptions on which the method is based are not satisfied in this case. As in practice this disagreement can be quite common, it may justify the little impact of the method observed in research literature after some initial popularity. The papers studying the transformation method were certainly of great value to improve the knowledge on distillation columns, but their use in practical problems requires a careful preliminary analysis. Some old "truths" must probably be re-evaluated to be put in their real dimension.

Searching for the best control structure for the studied column the behavior of each structure was simulated against feed flowrate (+/- 10%), temperature (+/- 1%; the system is highly sensitive to this disturbance) and composition (+/- 10%) disturbances. Controllers were tuned in such way as to get the best performance for each structure. Results in Figure 1 were obtained for the top of the column against the positive flowrate and composition disturbances, and the negative temperature disturbance. In addition to the controlled variables responses, the behavior of primary manipulated variables is also shown. Figure 2 displays similar results for the bottom of the column.

Considering the top (Figure 1), where composition control was more rigorous, it can be notice that there are close relations between the disturbances effects on the composition. The effect of 10% variation in feed flowrate was almost equivalent to -1% variation in feed temperature, while the effect of +10% variation in feed composition was less significant and in the opposite direction. The differences observed in the magnitude of the initial reflux flowrate have been attributed to the larger vapor amount required by the addition of the temperature control loop in the LV-cascade structure. But the fact that after 1% reduction in the feed temperature all the structures lead to the same reflux flowrate value still requires justification. There are large performance differences between the best (LV-cascade and [L/D,V/B]) and the worse (DV and LV) structures. Observing the light top mass fraction responses it can be verified that the best structures succeed in keeping the setpoint, with a minor advantage for the LV-cascade structure.
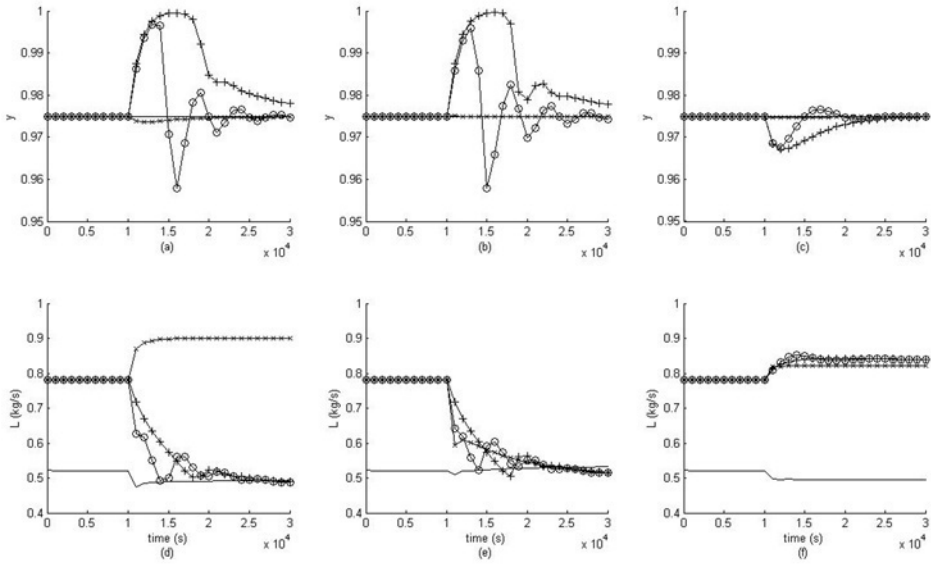
Figure 1. LV (+), DV (o), [L/D,V/B] (x) and LV-cascade (-) responses for feed flowrate (a, d), temperature (b, e) and composition (c, f) disturbances; sum of distillate light components mass fractions (a, b, c), reflux mass flowrate (d, e, f).
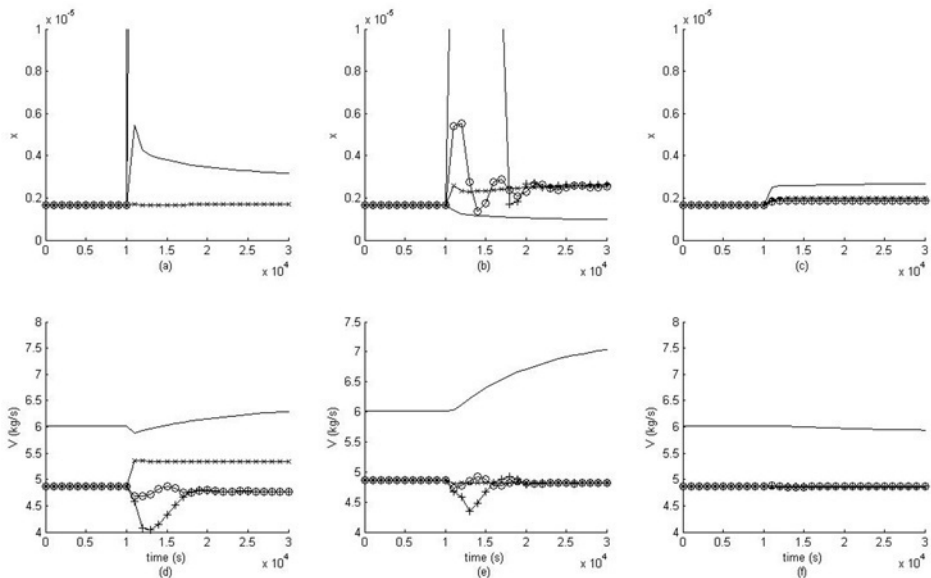
Figure 2. LV (+), DV(o), [L/D,V/B] (x) and LV-cascade (-) responses for feed flowrate (a, d), temperature (b, e) and composition (c, f) disturbances; sum of bottom's light components mass fractions (zoom) (a, b, c) and vapor mass flowrate coming from the reboiler (d, e, f).

The control efforts showed by changes in manipulate variables are almost equal, with a notable exception in the case of the (L/D, V/B) structure submitted to a disturbance in F, were the manipulated variable presents an inverse behavior. There is no simple explanation to this fact, but it is important to remind that the complex multicomponent feed, with small amounts of light component, could produce complex thermodynamic effects. Considering the bottom, where the tuning was less rigid, it can still be observed the best performance of LV-cascade and (L/D,V/B), but now it is not possible to distinguish any clear difference between them. It should be note that in this case disturbances effects are very small due to non-significant light components concentration.

3. Conclusions

The information obtained from simple static indexes was not sufficient to determine a best control structure. RGA results agreed with Shinskey (1984), apparently showing that the (L/D, V/B) structure has better decoupling characteristics than DV, and this than LV. Although there is not enough room for a detailed presentation, it is worth mentioning that setpoint top composition changes produced larger interaction effects in the structure (L/D, V/B) than in DV, in opposition to the previous statement. Moreover, the best performance during the simulation tests was obtained for the LV-cascade structure, whose index γ^* was the worst. It is interesting to note that, for the studied case, the positive characteristics of the "general" structure proposed by Skogestad (2007) were confirmed. The main conclusion that can be extracted form this work is that every effort should be made to have accurate dynamic simulators to aid the design of distillation columns control structures.

Acknowledgment

References

E. H. Bristol, 1966, On a New Measure of Interaction for Multivariável Process Control, IEEE Trans. Auto. Con., AC-11, pp.133-134.

P. Grossdidier, M. Morari, B. R. Holt, 1985, Closed-Loop Properties from Steady-State Gain Information, Ind. Eng. Chem. Fundam., 24, pp. 221-235.

K. E. Häggblom, K. V. Waller, 1988, Transformations and Consistency Relations for Distillation Control Structures, AIChE Journal, 34, 1634-1648.

F. G. Shinskey, 1984, Distillation Control, McGraw-Hill, New York.

S. Skogestad, 2007, The Dos and Don´ts of Distillation Column Control, Trans IChemE, Part A, 85(A1), pp. 13-23.

Development of Multivariate Statistical-Based Tools for Monitoring of Sour Water Unit

Douglas F. B. Lima[a], Fernando A. Zanella[a], Alexandre C. Teixeira[a], Luiz F. L. Luz Jr.[b], Carlos A. U. Gontarski[b], Enrico M. Gomes[b], Saulo H. Chiquitto[b], Marcelo K. Lenzi[b]

[a]*Refinaria Presidente Getúlio Vargas REPAR/PETROBRAS – Setor de Otimização
Rodovia do Xisto, BR476, Km 16, Araucária – PR, 83700-970, Brazil
e-mail: dfalleiros@petrobras.com.br*
[b]*Departamento de Engenharia Química, Universidade Federal do Paraná
Centro Politécnico, Jardim das Américas, Caixa Postal 19011, Curitiba – PR, 81531-980, Brazil*

Abstract

High production combined with high product quality and low environmental damage and energy consumption represents one of the main challenges of any chemical plant. In petrochemical plants, sour water treatment unit plays a key role regarding environmental issues, due to the necessity of ammonia and sulphur-based compounds removal. Although fundamental-based models can be successfully used for monitoring tasks, faster models based on multivariate statistics theory have been widely introduced in industrial sites. They present attractive features such as the capacity of handling a high amount of data, contain the process history and can be used for on-line and real-time process monitoring. In this study, multivariate statistical based methods were used to develop an alternative tool for an actual operating sour water treatment unit monitoring. Input variables collected from the process history were used to not only adequately predict the behavior of a given response variable, but also another potential use of this is to gain understanding of how and to which extent different process variables affect a specific response variable. Some issues like the length of predictive horizon, data structure, data selection and treatment during the development of the tool are also discussed. The obtained results indicate that multivariate techniques were able to monitor the unit variables behavior, representing a key alternative tool for sour water unit monitoring.

Keywords: sour water, multivariate statistics, process monitoring.

1. Introduction

High production combined with high product quality and low environmental damage and energy consumption represents one of the main challenges of any chemical plant. In petrochemical plants, sour water treatment unit plays a key role regarding environmental issues, due to the necessity of ammonia and sulphur-based compounds removal (MANDAL, 2007).

Ammonia and sulphur-based compounds, with minor amounts of other chemicals such as cyanides, are the main constituents of sour gas streams. The treatment

of sour gas represents a challenging issue (GAI *et al.*, 2008). One of the most common design strategies for sour water purification units is the use of two stripping columns. The first is used for sulphur-based (mainly H_2S) removal. The second column is used for ammonia removal (HARDISON, 1988). Further details regarding process design peculiarities, operating conditions and optimization issues can be found elsewhere (van HOORN, 2002).

Although fundamental-based models can be successfully used for monitoring tasks, faster models based on multivariate statistics theory have been widely introduced in industrial sites (HARRIS *et al.*, 1999). They present attractive features such as the capacity of handling a high amount of data, contain the process history and can be used for on-line and real-time process monitoring. According to MILETIC *et al.* (2004) multivariate statistics have been used for a wide range of industrial applications such as a sulfite pulp digester plant, desulfurization process, among other.

Several multivariate statistical methods are reported in literature, for example, principal component analysis (PCA), partial least squares (PLS). Further details concerning the methods, such as advantages, formulation, and application, can be found elsewhere. BRERETON (2003), OTTO (1999), JOHNSON & WICHERN (2002).

In this study, multivariate statistical methods were used to develop an alternative tool for an actual operating sour water treatment unit monitoring. Input variables collected from the process history were used to not only adequately predict the behavior of a given response variable, but also another potential use of this is to gain understanding of how and to which extent different process variables affect a specific response variable. Some issues like the length of predictive horizon, data structure, data selection and treatment during the development of the tool are also discussed. The obtained results indicate that multivariate techniques were able to monitor the unit variables behavior, representing a key alternative tool for sour water unit monitoring.

2. Methodology

Data from REPAR process plant information (PI) of three complete years were collected and analyzed. The analysis comprised statistical tests already reported in the literature (HIMMEMLBLAU, 1970) to obtain data sets which represent process steady state operation. After a steady state operation period was found, the values as they are were used in the analysis and no outlier test was conducted. Linear regression analysis was used in order to obtain different empirical models in order to predict important process variables. Further details of the linear regression analysis and parameter estimation methods can be found in the literature. (JOHNSON & WICHERN (2002)). The models have a basic structure given by:

(dependent variable) = constant + (coefficient · independent variable)$_1$ + … + (coefficient · independent variable)$_N$.

Temperature measurements were obtained with a YTA-110 YOKOGAWA temperature transmitter equipped with a thermocouple of type K. Pressure measurements were obtained with EJA-430A YOKOGAWA pressure transmitter. Finally flow rates were obtained with 3051 EMERSON device. This work focused the first stripping column of the sour water unit of REPAR. For each case studied, the model which generated the predicted values is presented. Different independent variables were tested and the best obtained results are reported here. Plots of predicted values of the desired variable was against the real (observed) values are also presented. These plots were used to obtain the presented correlation coefficient. Results are

normalized and units are omitted for presentation. It must be stressed that these models are used to check the obtained value of the process sensor. Data horizon of 3 years was used. Data were treated using moving average data filter in order to eliminate mean less information. After filtering, operation steady states were identified/recognized by exploring mean and variance of collected data in a given horizon, which could range from one to five hours, using techniques reported in the literature (MaCAULEY, 2003). The linear least square technique used is a straightforward technique, representing an important tool for seeking correlation among process variables.

3. Results and Discussion

The first model can be used for prediction of the pressure at the top of the stripping column (dependent variable). The model equation is given by a linear combination of the variables presented in Table 01.

Table 01 – Variables of model 01

Independent Variable	Estimated Coefficient
Steam temperature at the exit of the column reboiler.	0.01436
Temperature at the top of the column	0.002021
Column Pressure Drop	-0.18614
Ambient Temperature	0.008012
Constant	6.209758

Figure 01 presents a plot of predicted versus real values. The correlation coefficient between predicted and real values of the pressure is 0.85. It can be seen from the picture that model predictions are reasonable for values over 0.5 up to 1. The value of the correlation bellow 0.85 probably occurs due to the experimental below 0.5 which were not omitted in the analysis.
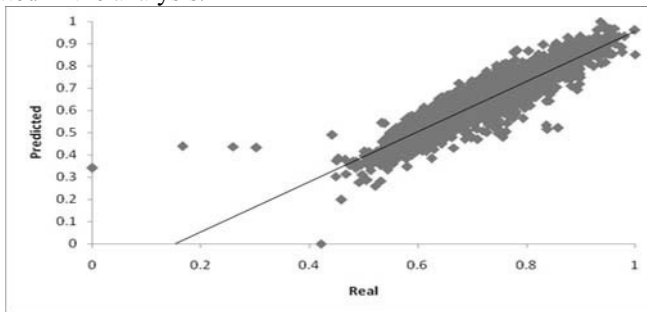
Figure 01 – Predicted versus real values of the dependent variable – model 01

It must be emphasized that not outlier test was conducted and better prediction can be obtained by either increasing the number of model parameters or changing the independent variables. Although linear, the used model did not took into account variables interactions as they would only increase the amount of parameters and did not significantly improve the correlation coefficient, showing that from a statistical point of view they are not relevant.

As a consequence, a second model can be used for prediction of the pressure at the top of the stripping column, leading to better predictions when compared to the first model. Different independent variables were chosen for this second model. The model equation is given by a linear combination of the variables presented in Table 02. The obtained correlation coefficient was of 0.85. Although the correlation coefficient was

roughly the same, this model is better because the range from 0 to 1 of predicted values lie over the straight line, differently from the previous model that values ranging from 0 to 0.5 cannot be adequately predicted.

Table 02 – Variables of model 02

Independent Variable	Estimated Coefficient
Condenser pressure	0.280782
Ambient Temperature	0.007457
Column feed stream temperature	0.004443
Temperature of cold inlet heat exchanger fluid	0.002483
Constant	5.60264
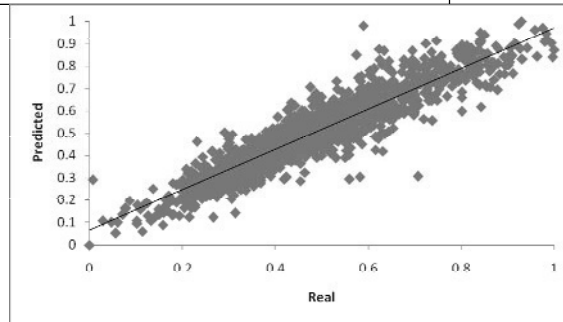
Figure 02 – Predicted versus real values of the dependent variable – model 02

Another important variable is the amount of live steam need for column operation. Our model presented a correlation coefficient of 0.83, which is a reasonable value. Table 03 presents the variables used in model. Predicted against real values are shown in Figure 03. The model present the data tendency as predicted values increase as does observed values. On the other hand, the chart presents three cluster data regions. This probably happened due to either a nonlinear relation among these variables or a strong interaction among the variables.

Table 03 – Variables of model 03

Independent Variable	Estimated Coefficient
Pressure at the top of the column	1.997464
Column feed rate	0.083669
Temperature at the top of the column	0.009882
Column Pressure Drop	2.186367
Constant	-17.412
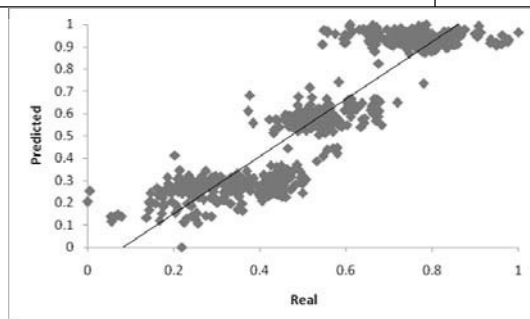
Figure 03 – Predicted versus real values of the dependent variable – model 04

The models previously presented are used to predict variables which are continuously measured. However, one of the aims of this work is the development of models to accurately predicted variables which are not continuously measured such as the NH_3 concentration at the bottom of the column. Table 04 presents the independent variables used for model derivation.

Table 04 – Variables of model 04

Independent Variable	Estimated Coefficient
NH_3 concentration in the column feed stream	0.927587
Column feed rate	0.083669
(Column feed rate)·(0.001·(NH_3 concentration in the column feed stream)) / (column bottom exit rate)	-200.267
Constant	634.2305

A poor correlation coefficient has been achieved 0.75. However it should be stressed that outlier data was not removed in order to keep all possible real situation observed during plat steady state operation. Besides, only a linear model on the parameters was worked. The use of nonlinear models on the parameters may improve the correlation coefficient. On the other hand, it can be seen that the model predicted values present the same tendency of the real data. Figure 04 presents predicted values against real data.
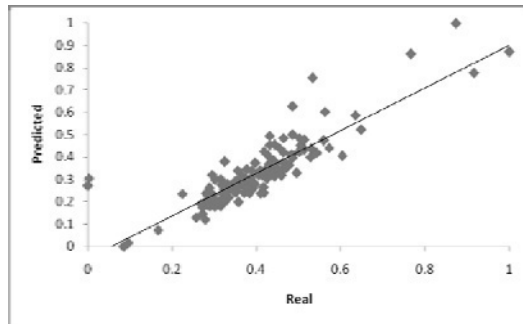
Figure 04 – Predicted versus real values of the dependent variable – model 04

Finally, the column H_2S recovery could be successfully modeled by a linear model. Table 05 presents the independent variables and the estimated coefficients.

Table 05 – Variables of model 05

Independent Variable	Estimated Coefficient
NH_3 concentration at the top of the column	33.3646
(NH_3 concentration at the top of the column)^2	-4.16765
Temperature at the top of the column	-0.15596
Temperature variation at the column reboiler	-0.36381
Temperature variation at the column	-0.1567
Constant	-38.5063

This model could successfully describe the column H_2S recovery as a correlation coefficient of roughly 0.94 has been achieved. Besides, the plotted values are randomly distributed on both sides of the line indicating that no bias is present. Also, the data is not clustered in two or more regions which might also lead to a good correlation coefficient and all range from 0 to 1 can be predicted.
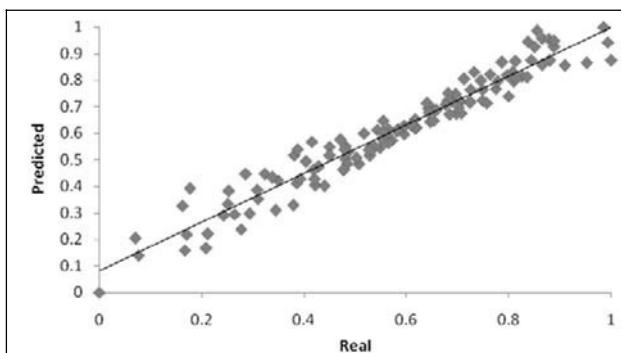
Figure 05 – Predicted versus real values of the dependent variable – model 05

4. Conclusions

The models were able to successfully describe different process steady states. These models represent and alternative tool for process monitoring in case of sensor failure. Although linear, the models can adequately describe the actual process values, being an important tool mainly for prediction of intermittent data which are not continuously measured such as chemical concentrations, or reduce lab analysis. It can also be used for sensor measurement checking.

5. Acknowledgements

References

R.G. BRERETON, 2003, Chemometrics: Data Analysis for the Laboratory and Chemical Plant. John Wiley & Sons: London.

H. GAI; JIANG,Y.; QUIAN,Y.; KRASLAWSKI,A., 2008, Conceptual Design and Retrofitting of the Coal-Gasification Waste Water Treatment Process. Chemical Engineering Journal. 138, 84-94.

L.C. HARDISON, 1988, Removal of Hydrogen Sulfide From Sour Water. US. PATENT 4.784.775.

T.J. HARRIS; 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. 9, 01-17.

D.M. HIMMELBLAU, 1970, Process Analysis by Statistical Methods. 1ST Ed. John Wiley & Sons: New York.

R.A. JOHNSON; WICHERN,D.W., 2002, Applied Multivariate Statistical Analysis. 5TH Ed. Prentice-Hall: Upper Saddle River.

K.K. MANDALL, 2007, Optimize Your Gas Concentration Unit Operations. Hydrocarbon Processing. 86, 51-56.

K.B. MacCAULEY, 2003, Analysis of Process Data. Internal Report. Queen's University.

I. MILETIC; QUINN,S.; DUDZIC,M.; VACULIK,V.; CHAMPAGNE,M., 2004, An Industrial Perspective on Implementing On-Line Applications of Multivariate Statistics. Journal of Process Control. 14, 821-836.

M.OTTO, 1999, Chemometrics: Statistics and Computer Application In Analytical Chemistry. 1ST Ed. Wiley-VHC: Weinheim.

E. Van HOORN, 2002, Basics of Sour Water Stripping. Amine Experts, 411-444.

Operational Strategy for Water Supply in a Petrochemical Plant. Steady-State and Dynamic Approaches

Rita M. B. Alves, Antonio E. Bresciani, William S. Maejima, Claudio A. O. Nascimento

LSCP/CESQ - Department of Chemical Engineering, Polytechnic School - University of São Paulo, Av. Prof. Luciano Gualberto, n. 380, tv. 3, CEP 05508-900 São Paulo, SP, Brazil.

Abstract

The aim of this work is to present a mathematical model developed in order to simulate several operating scenarios involving water supply to a vapor generation unit in a petrochemical plant. The operational strategy suggested involves steady state and dynamic considerations. The case study is a petrochemical plant that uses water from two distinct sources with different qualities and prices. Thus, two case studies were carried out: 1. definition of the optimal proportion between both sources of water in case of supply stability: a model based on the minimum total cost per cubic meter of useful water was developed; 2. definition of the best operational strategy in case of failure in the main water source: a dynamic model was developed representing the flow rates as a time-dependent mathematical function in order to evaluate possible disturbances and all the operating conditions of interests, including the risk of total lack of water for vapor generation, which would cause the whole plant to shut down. The results show that it is possible to preserve operational continuity and stability, with lower water consumption and wastewater generation.

Keywords: Water supply, Optimization, Steady-state model, Dynamic Modeling.

1. Introduction

Water is an essential resource in industrial activities. In most process industries water is vital for many operations and is utilized for different purposes. In many regions with high concentration of industrial activities, or those close to urban centers, the operation of industrial plants is limited by the reduction in the capacity of water sources to supply demand, by increasing restrictions in the discharge of effluents into the environment and by competition with urban consumption. This situation has led to high costs of water supply and wastewater treatment (Goldblatt et al., 1993), resulting in a powerful economic motivation to rationalize water use, and an increasing the need for adequate water management and wastewater minimization, especially in the petrochemical industry, due to the large rates of consumption involved. Water and wastewater minimization techniques have been widely researched, developed and applied to process industries. Bagajewicz (2000) presented a review of recent design procedures for water networks in refineries and process plants. The basic concept of water minimization is to maximize water reuse and to identify regeneration opportunities (Wang and Smith, 1994). In general, there are four approaches to water minimization (Zver and Glavic, 2005): process changes, reuse, regeneration reuse and regeneration recycling. On the

other hand, process engineers seek simpler alternatives for reducing water demand in their plants. One strategy consists of evaluating individual process and utility units to reduce their inherent water requirements. Examples include studying operational strategies in order to guarantee water optimization in steady-state and transient scenarios.

Industrial water management and optimization require, in many cases, tailor made solutions. Thus, the present work presents a mathematical model developed for simulating several operating scenarios involving the water supply to a vapor generation unit in a petrochemical plant. The operational strategy suggested involves steady-state and dynamic considerations. The results show that it is possible to preserve the operational continuity and stability, with lower water consumption and wastewater generation at reduced cost.

2. Problem Statement

The case study is a petrochemical plant located in a region with a high concentration of industrial activities and close to an urban agglomeration. This industry uses water from two distinct sources: A and B, tap water and industrial water, respectively. Water from source A presents good quality, however at a high price and with low supply reliability. On the other hand, water from source B has, conversely, low quality, low price, and high supply reliability. Part of the water fed to the plant is treated in a demineralization unit in order to remove dissolved mineral matter that causes the water to be hard, thus making it proper to feed the vapor generation unit.

Figure 1 shows the studied scheme of water supply to the demineralization process, which consists of three sets of cationic and anionic resin beds, M_1, M_2 and M_3. Eventually, the resin beds become saturated, and at this point they stop softening the water. It is then time to regenerate it. It is considered as a semi-batch operation. Two batteries operate simultaneously while another one is being regenerated. Thus, it is possible to provide the required water flow rate to the vapor generation unit. The operating cycle length is proportional to water hardness.
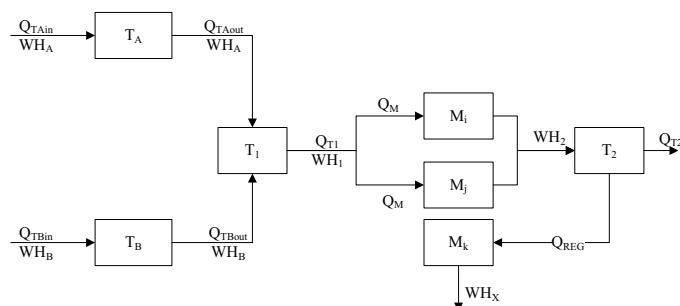
Figure 1 - Demineralization unit

The level of the tank that contains water from source A, T_A, is controlled at 90% and the tank with water from source B, T_B, is aligned to provide water to the system. In order to reduce costs, the studied plant used to feed the demineralization unit with a mixture from both sources in a pre-defined proportion; in general, 15-20% from source B. Assuming a fixed regeneration time and fixed flow rate of treated water used for regeneration of the resin beds, a greater operating cycle length is expected whenever the unit uses more volume of water from source A, which presents lower hardness than

water from source B. Thus, two main case studies were carried out: 1. definition of the optimal proportion between both sources of water in case of supply stability – the steady-state model; 2. definition of the best operational strategy in case of failure in the supply of water from source A. Matlab version 6.5 was used to develop the models.

3. Modeling

Figure 2 shows the algorithm developed in order to model the studied system.
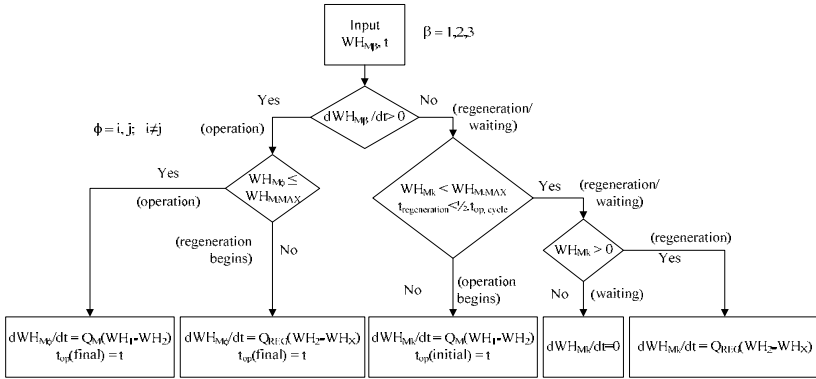
Figure 2 - Algorithm developed to model the system

3.1. Steady-State Model

The steady-state model aims to establish the optimal proportion between both sources of water in case of supply stability, taking into consideration economical aspects. The developed model was based on the minimum total cost per cubic meter of useful water. In order to carry out the calculations, the following conditions were assumed:

- the hardness of water from sources A and B are $WH_A = 30$ ppm and $WH_B = 100$ ppm, respectively;
- the maximum hardness removed by the demineralization unit, WH_{M_max}, is 48600 g/operating cycle. This value was evaluated through the Eq. (1) by assuming the usual proportion of water from source B ($x_B = 0.15$) and 1200 m³/operating cycle;

$$WH_{M,max} = V_{Total} . [(x_A . WH_A) + (x_B . WH_B)]$$
(1)

- the total flow rate of water treated in the demineralization unit, V_{Total} (m³/operating cycle), was evaluated by Eq. (2);

$$V_{Total} = 48600 / \{(x_A . 30) + [(1 - x_A).100]\} \qquad x_A = (0...1)$$
(2)

- The number of operating cycles per day, Ncycles, that the demineralization unit is able to run was calculated by Eq. (3);

$$Ncycles = \frac{Q_{T2} \cdot 24}{V_{Total} - Q_{REG}} \qquad \textit{(operating cycles/day)}$$
(3)

where Q_{T2} (≥ 110 m³/h) corresponds to the treated water flow rate used to feed the vapor generation unit and Q_{REG} (240 m³/operating cycle) is the treated water flow rate used to regenerate the resin bed.

Eq. (4) was used to evaluate the total cost per hour of the process, *Cost 1*, including the costs for water acquisition, C_A and C_B, and the operational costs, C_{op}.

$$Costs\ 1\ (US\$/h) = Q_{TA} \cdot C_A + Q_{TB} \cdot C_B + C_{op} \cdot f \tag{4}$$

where C_A = US\$ 3.00/m^3 ; C_B = US\$0.70/m3; C_{op} = US\$ 1000.00/operating cycle *and f= (Ncycles/24)*. Eq. (5) shows the total costs per cubic meter of treated water.

$$Costs\ 2\ (US\$/\ m^3) = \frac{Costs\ 1}{(V_{Total} - Q_{REG})f} \tag{5}$$

Table 1 shows the total cost as a function of the proportion of water from source A.

Table 1 – Results of the steady-state model

Source A (%)	0	10	20	30	40	50	60	70	80	85	90	100
Ncycles	10.73	9.34	8.12	7.04	6.07	5.2	4.42	3.7	3.05	2.75	2.46	1.91
Costs (US\$/m3)	5.45	5.26	5.09	4.94	4.81	4.69	4.59	4.49	4.40	4.36	4.32	4.25

As the proportion of water from source A increases, the number of operating cycles per day and the total cost decreases. This result shows that the costs of water acquisition are compensated by the operational costs, i.e., the better the water quality, the higher the length of the operating cycle and the lower the treatment costs. Thus, the optimal operating condition entails using water solely from source A, despite its higher price. Operation with 100% of water from source A, compared to the normal operation, would lead to the following advantages: reduction in the total useful water cost, equivalent to about US\$ 110 thousand per year; and a reduction of 73365 m^3/year in the total water consumption and wastewater generated in the process.

3.2. Dynamic Model

A model was developed in order to dynamically simulate failure scenarios in the supply of water from source A, a somewhat frequent occurrence. The objective was to evaluate the process in the hypothesis of reduced flow rate or total lack of water from source A, allowing fluctuation in the level of tanks T_A and T_2, but keeping a constant water flow rate to the vapor generation unit in order to guarantee operational continuity and stability. The flow rates are represented as a time-dependent mathematical function in order to evaluate possible disturbances and all the operational conditions of interest.

In order to dynamically represent the system, a set of material balances was carried out:

$$\frac{dV_{TA}}{dt} = Q_{TAin} - Q_{TAout} \tag{6}$$

$$\frac{dV_{TB}}{dt} = Q_{TBin} - Q_{TBout} \tag{7}$$

$$\frac{dV_{T1}}{dt} = Q_{Taout} + Q_{TBout} - 2Q_M \tag{8}$$

$$\frac{dWH_1}{dt} = \frac{Q_{TAout} \cdot WH_A}{V_{T1}} + \frac{Q_{TBout} \cdot WH_B}{V_{T1}} - \frac{2Q_M \cdot WH_1}{V_{T1}} \tag{9}$$

$$\frac{dWH_{M\phi}}{dt} = Q_M \cdot (WH_1 - WH_2) \qquad \phi = i \ and \ j \qquad operating \ cycle \tag{10}$$

$$\frac{dWH_{Mk}}{dt} = Q_{REG} \cdot (WH_2 - WH_X) \qquad\qquad resin \ regeneration \tag{11}$$

$$\frac{dV_{T2}}{dt} = 2Q_M - Q_{REG} - Q_{T2} \tag{12}$$

The dynamic study was carried out assuming the optimal operating condition as defined in the steady-state study, i.e., only water from source A was used instead of a mixture of both sources, considered by this industry as the standard operating condition. The optimal operating conditions were evaluated by minimization of an objective function defined in terms of total cost according to Eqs. (4) and (5).

4. Results and Discussion

Different scenarios involving the failure of water supply from source A were evaluated: flow rate reductions of 25%, 50%, 75% and 100%. The minimum cost was obtained at tank T_A level equal to 10% for all scenarios studied. The observed differences in the costs calculated as a function of the tank level were not significant.

For the dynamic simulation, the following conditions were assumed:

- after total failure of the water supply from source A ($Q_{TAin} = 0$), the level of the tank T_A decreases until it reaches a minimum of 10%.
- to avoid entire plant shutdowns, some action should be taken before the level of the tank T_2 reaches a minimum of 30%.
- fixed regeneration time and fixed flow rate of treated water used for regeneration;
- the regeneration time, t_{REG} cannot be greater than the operating cycle time, $t_{op-cycle}$;
- water hardness of the regeneration process output stream, WH_X, is constant and equal to 1000ppm.

4.1. Case Study 1 :

In this case, the cause of the failure was fixed and the flow rate of A, $Q_{TAin,}$ came back to its original value 11.4 hours after the failure takes place. Water from source B was not used. Figure 3 shows the results. Just after the water supply is normalized, the three demineralization unit batteries start to run simultaneously. Since the water flow rate to each bed set is lower than that fed during normal operation, the beds take longer to become saturated and the regeneration process carries out. However, it is interrupted and the three beds are required to operate in order to guarantee the treated water supply to the vapor generation unit. Over time, the whole operation tends to achieve the steady-state regime.

4.2. Case Study 2:

In this scenario, water from source B was required to feed the system. Water from source A was not used. Figure 4a shows the results. The system behavior is similar to the case study 1. A higher number of operating cycles is observed since water from source B presents greater hardness than water from source A and the process takes longer to achieve a steady-state operation. The system presents a great oscillation which can cause lost of stability in the entire plant or reduction in the quality of treated water. This means that a quick action can avoid operational issues. As an example, tank T_B

was aligned to provide the required water flow rate to stabilize the system and restore the original level of tank T_2. Figure 4b shows this result.
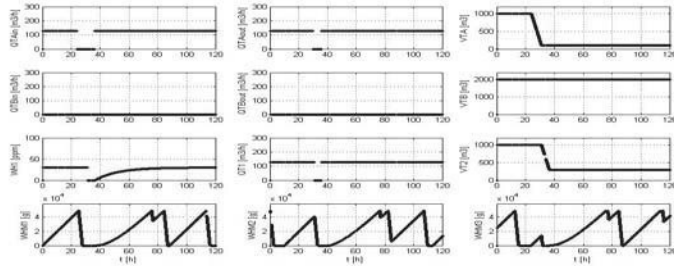
Figure 3 - Case study 1 simulation results

(a)

(b)
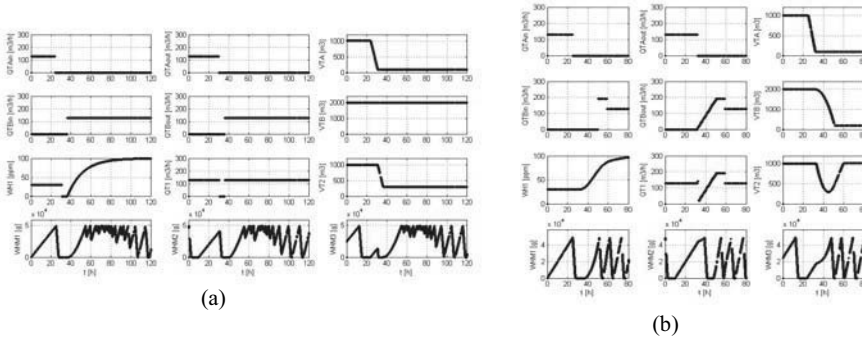
Figure 4 - Case study 2 simulation results

5. Conclusion

The simulation of steady-state scenario shows that is possible to optimize the system with reduction in the total useful water cost as well in the water consumption and wastewater generated in the process.

The developed model was able to represent the behavior of the transient regime due to disturbances in the water supply. In emergency situations, considering safer operating conditions rather than minimum cost as a deciding factor, in order to take appropriate actions, is recommended. Thus the system should be able to return to the steady-state condition keeping the quality of treated water and avoiding the plant shutdowns.

Acknowledgments

References

M. E. Goldblatt, K. S. Eble and J. E. Feathers (1993) Zero Discharge: What, Why and How, Chemical Engineering Progress, 89(4) 22-27.

M. Bagajewicz (2000) A review of recent design procedures for water networks in refineries and process plants, Computers and Chemical Engineering, 24(9-10): 2093-2113

Y. P. Wang and R. Smith (1994) Wastewater minimisation, Chemical Engineering Science, 49(7): 981-1006

L. Žbontar Zver and P. Glavic (2005) Water minimization in process industries: case study in beet sugar plant, Resources, Conservation and Recycling, 43(2): 133-145

Best Feature Selection for Texture Classification

Daeyoun Kim[a], J. Jay Liu[b], Chonghun Han[a]

aSchool of Chemical and Biological Engineering, Seoul National University, San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, South Korea
bSamsung Electronics, Myeongam-ri 200, Tangjeong-myeon, Asan, Chungchengnam-do 336-840, South Korea

Abstract

Texture analysis techniques enable to determine the quality of product surfaces measured by image sensors. In previous works, wavelet texture analysis based on the conventional wavelet transform and wavelet packets have been recognized as the most successful technique for classifying steel quality. In this work, we propose a texture classification strategy based on a best feature selection method, which improves classification accuracy. Our proposed methodology has been applied and validated in the classification of the surface quality of rolled steel sheets.

Keywords: Wavelet texture analysis, texture classification, feature selection, product surfaces, best features selection.

1. Introduction

The visual quality of a product surface must be controlled or maintained for most product manufacturers.(Liu *et al.*, 2007) The surface quality of steel sheets is important to automakers since it can significantly affect the coating quality. Some visual characteristics of product surfaces can be measured by digital image sensors. Texture analysis enables to determine the quality of product surfaces by using image sensors for measurement. Texture can be defined as an attribute representing the spatial arrangement of the gray levels of the pixels in a region of a digital image.(Geraci, 1990) Texture analysis has been successively applied to determine quality of product surfaces.(Chang *et al.*, 1993) Previous work(Bharati *et al.*, 2004) found that wavelet texture analysis offered the best performance in texture classification. Liu *et al.*(Liu, *et al.*, 2007) proposed an advanced methodology based on wavelet packets for characterizing steel surfaces.(Bharati, *et al.*, 2004) As wavelet packets use a large number of texture signatures, it has greater potential to include discriminative features for classification. These over-complete wavelet packets also impart increased computational costs and unnecessary features in the classification process. In order to use wavelet packets as a feature extraction method, the significant features for classification from full wavelet packets need to be selected.

This work proposes a texture classification methodology that selects the best feature from full wavelet packets. The proposed methodology was implemented and validated in a steel surface case. The results of the proposed methodology show performance improvements as compared to conventional wavelet transform and the full wavelet packets.

2. Theory for Wavelet Texture Analysis

2.1. Conventional Wavelet Transform

For multi-resolution representations, using wavelets is well established and successfully applied in signal and image processing.(Livens *et al.*, 1997) The principle is essentially filtering a signal through a pair of filters—a low-pass filter (*h*) and a high-pass filter (*g*)—and down-sampling the filtered signals by two (i.e., dropping every other sample), respectively.(Chang *et al.*, 1992) A signal *f(x)* is decomposed into its components through shifting and dilating of a prototype function *ψ(x)*.(Rajpoot, 2002) In this algorithm, a detail section is extracted at each decomposition level from the approximation section of the previous level, and the number of detail sections is reduced to half the number of the previous level. This is termed a conventional wavelet transform.(Chang*, et al.*, 1992; Liu*, et al.*, 2007)

2.2. Wavelet Packets

The conventional wavelet transform has a set of frequency channels comprising narrower bandwidths in the lower frequency region. In many practical cases, information on natural textures is not found in the lower frequency region, but rather in the middle or higher frequency region. These are termed quasi-periodic signals.(Chang*, et al.*, 1993; Liu*, et al.*, 2007) To analyze quasi-periodic signals, a generalized form of wavelet termed a wavelet packet is needed.(Rajpoot, 2002) A conventional wavelet transform is implemented through an iterative decomposition of approximation coefficients, using a two-channel filter bank. On the other hand, wavelet packets use all the coefficients to decompose, resulting in an equal frequency bandwidth.(Chang*, et al.*, 1993; Liu*, et al.*, 2007)

2.3. Wavelet Texture Analysis

Wavelet texture analysis is a very powerful method as compared to other texture analysis methods.(Rajpoot, 2002; Liu*, et al.*, 2007) To reflect scale-dependent characteristics, textural features can be extracted from each sub-image.(Livens*, et al.*, 1997) Each sub-image from wavelet decomposition is represented as D(J, I), where J is the depths of level of decomposition and I is the number of packets in the depths. The energy of the sub-image can be defined as follows:

$$E_{D(J,I)} = \left\| D(J,I) \right\|_F^2 \tag{1}$$

The energy divided by the number of pixels equals the averaged power (normalized energy). The most popular wavelet textural feature, the wavelet energy signature, is a vector organized by the energies of all the sub-images.(Kim *et al.;* Kim *et al.*, 2007; Liu*, et al.*, 2007)

3. Texture Classification Strategy Based on Best Feature Selection

The proposed texture classification methodology based on the best feature selection has four steps: feature extraction, feature selection, feature reduction, and classification.

3.1. Feature Extraction Step

The extraction of important textural information from images is essential for texture analysis. Bharati *et al.*(Bharati*, et al.*, 2004) compared these methods and concluded that the transform-based method—particularly the wavelet texture analysis—offered the best performance in terms of classifying steel textures. For texture classification, we selected a wavelet texture analysis based on wavelet packet bases, since this has been shown to perform better than other methods.(Liu*, et al.*, 2007)

3.2. Feature Selection Step

3.2.1. Criterion for feature selection

A criterion is needed to select the best features from wavelet packets. For this study, an index that represents class separation or distances among classes can be used.(Saito *et al.*, 1995) Fisher's criterion provides reliable class separation among the different bases. Let the between-class scatter matrix be S_B, and the within-class scatter matrix be S_W. Then, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns, which maximizes the ratio of the between-class scatter to the within-class scatter.(Duda *et al.*, 1973)

$$J(W_{opt}) = \frac{\left| W^T S_B W \right|}{\left| W^T S_W W \right|}. \qquad (2)$$

The objective value, denoted as $J(W_{opt})$, is termed Fisher's criterion. Fisher's criterion obtained from Eq. (2) can be used as an approximated classification performance measure for different datasets. Among the different feature vectors, a well-classifiable one will have a high Fisher's criterion value. We utilized these characteristics to measure a discriminative basis extracted from wavelet packets for the texture of images.(Rajpoot, 2002; Liu, *et al.*, 2007)

3.2.2. Simple searching method

We used a top-down searching method for feature selection. Feature selection was performed using a simple rule, which has already been used for other pattern recognition problems.(Kumar *et al.*, 2001) This simple agglomerative rule does not require recursive computation for sorting discriminations of features.

C_{opt} : best features consist of only selected wavelet energy signatures

C_{JI} : wavelet energy signatures at level J; the number of packets is I

$\overline{C_{JI}}$: decomposed wavelet energy signatures of C_{JI}

$\overline{\overline{C_{JI}}}$: decomposed wavelet energy signatures of $\overline{C_{JI}}$

As one moves from the upper level to the lower level—and from lower packet numbers to higher packet numbers in each level—the best features are updated by using simple selection rules as follows:

if $J(C_{JI} \cup C_{opt}) > J(\overline{C_{JI}} \cup C_{opt})$, then C_{opt} includes undecomposed

signatures C_{JI} , and

if $J(C_{JI} \cup C_{opt}) \leq J(\overline{C_{JI}} \cup C_{opt})$, then C_{opt} includes decomposed

signatures $\overline{C_{JI}}$,

where J is Fisher's criterion. Feature selection is performed from the upper level (level 1) to the lower level. Only the decomposed bases have higher criteria than the present levels, to be included in the best features. Otherwise, undecomposed bases were included.

Top-down searching requires an assumption for pruning appropriately.

$$\text{If } J(C_{JI} \cup C_{opt}) > J(\overline{C_{JI}} \cup C_{opt}), \text{ then } J(C_{JI} \cup C_{opt}) > J(\overline{\overline{C_{JI}}} \cup C_{opt}).$$

3.3. Feature Reduction Step

Further dimension reduction can be achieved by using projection methods. In this study, we selected Fisher's discriminant projection for maximizing discrimination among textural classes.

Figure 1. Steel Surface Sheets(Kim*, et al.;* Liu*, et al.*, 2007)

(a) Excellent   (b) Good   (c) Medium   (d) Poor
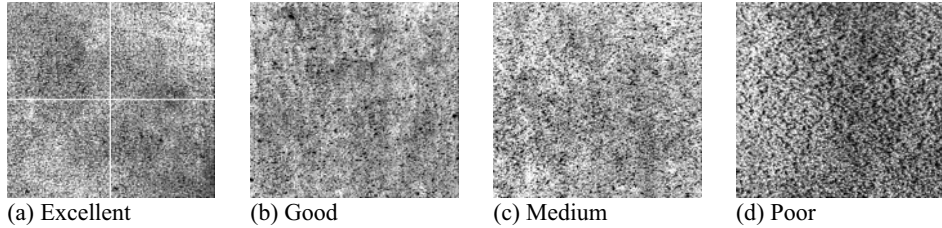
3.4. Texture Classification Step

Classification algorithms can be differentiated according to their features. In this study, we used K-nearest neighbors, a simple supervised classification algorithm, to classify the features.

4. Application to Classification of Steel Surfaces

4.1. Steel Surfaces

To validate our proposed methodology, we compared proposed best features with previous conventional and full-wavelet packets in the classification of steel-surface quality. The quality of each steel surface was determined as excellent, good, medium, or poor, as shown in Fig. 1. For this study, 35 images of steel surfaces were used. The example of bad surface quality contains deep pits that have joined to form craters. The number of samples within each class was 8 (excellent), 9 (good), 6 (medium), and 12 (poor). To obtain greater consistency, each original image was divided into four sub-images (see Fig. 1a). Therefore, the new image set contained 140 images.

4.2. Wavelet Texture Analysis

We applied a two-dimensional wavelet packet transform to all 140 images, using order-one Coiflet wavelet filters up to level 4. The maximum decomposition level used for all the three methods (best features, conventional wavelet transform, and wavelet packets) was four. The maximum decomposition level used for the wavelet transforms was selected in accordance with certain guidelines.(Chang*, et al.*, 1993) For feature extraction, each sub-image was converted to a one-scalar value.

4.3. Best Feature Selection

Conventional wavelet transform and full wavelet packets do not require feature selection and thus proceed to the next feature-reduction step. As explained earlier, best features were selected according to a top-down searching rule.

4.4. Quality Classification

Fisher's discriminant analysis maximizes discrimination in each class. Following Fisher's projection, we equally applied a three-nearest neighbor classification for all bases. The misclassification rate was estimated by using a "leave one out" cross-validation.

5. Results of Case Study

Figure 2 shows the feature reduction by Fisher's discriminant plot. In Fig. 2, level 4 wavelet packets and conventional wavelet transform have some linearly inseparable points. The good classes and excellent classes are not perfectly separable in wavelet packets. On the other hand, Fig. 1(c) shows that each class in the best features is perfectly separated.

(a)   (b)

(c)
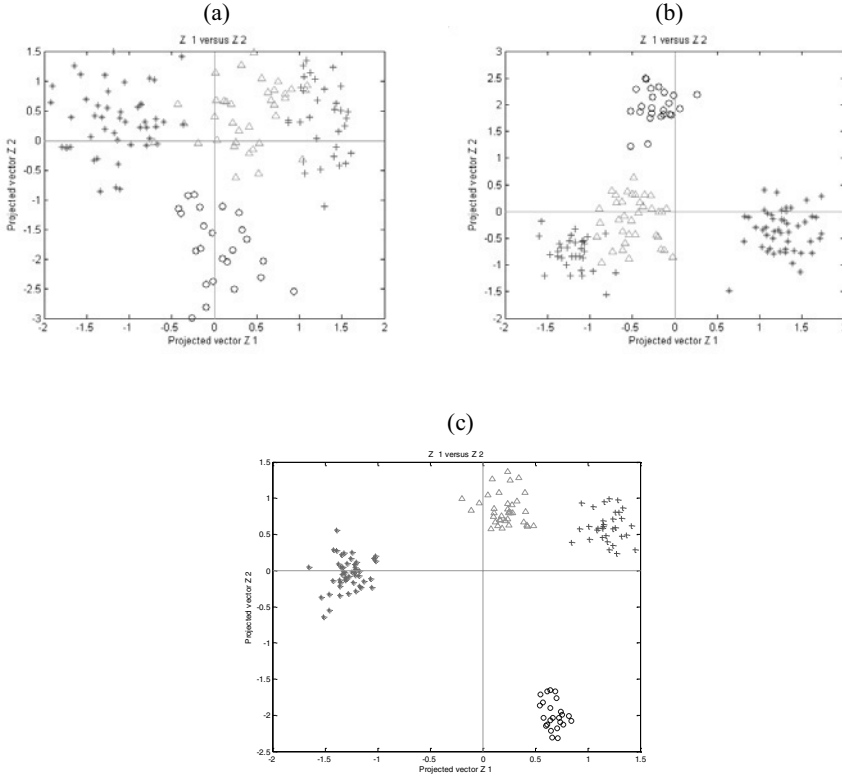
Figure 2. Fisher's Discriminant Plots.

(a) Conventional Wavelet Transform; (b) Full-Wavelet Packets; (c) Proposed Best Features.
Class labels are +: excellent; △: good; ○: medium; *: poor.

Table 1 shows the same results as Fig. 2 in a quantitative manner. The proposed approach achieves the best classification performance among the three methodologies (wavelet packets, conventional wavelet transform, and proposed best features). Figure 2 and Table 1 show that selected features from top-down searching suitably selected and improved classification performance.

Table 1. Classification Errors of Steel Surfaces

	Conventional Wavelet Transform	Wavelet Packets	Proposed
Excellent	5.9%	0.9%	0%
Good	24.1%	2.9%	0%
Medium	0%	0%	0%
Poor	1.1%	0%	0%

6. Concluding Remarks

A previous work employed wavelet texture analysis based on full wavelet packets or conventional wavelet transform for texture classification. These two approaches have some disadvantages with regard to such classification. Wavelet packets include insignificant packets, which lead to misclassification. Conventional wavelet transform does not provide sufficient packets to represent whole information. In this study, a texture classification methodology, based on best feature selection, was proposed and validated using steel-surface sheets. We compared the performance of the proposed method with that of full wavelet packets and conventional wavelet transform. This work proved that the proposed method is more appropriate for characterizing steel surface quality than both the conventional wavelet transform and full wavelet packets.

References

M. H. Bharati, J. J. Liu, and J. F. MacGregor, 2004, Image texture analysis: methods and comparisons: Chemometrics and Intelligent Laboratory Systems, 72, 57-71.

T. Chang, and C. C. J. Kuo, 1992, A wavelet transform approach for hierarchical texture segmentation, 812, 816-820.

T. Chang, and C. C. J. Kuo, 1993, Texture analysis and classification with tree-structured wavelet transform: Image Processing, IEEE Transactions on, 2, 429-441.

R. O. Duda, and P. E. Hart, 1973, Pattern Classification and Scene Analysis: Wiley New York.

A. K. Geraci, 1990, IEEE Standard Glossary of Image Processing and Pattern Recognition Terminology: Institute of Electrical and Electronics Engineers (IEEE) Std, 610.614-1990.

D. Kim, J. Liu, and C. Han, 2007, Comparison of features in classifying steel surface quality, ESCAPE 18proceedings.

D. Kim, J. Liu, and C. Han, 2007, Quality monitoring of steel surface using wavelet packet transfrom, ICCAS2007 proceedings, 861-864.

S. Kumar, J. Ghosh, and M. M. Crawford, 2001, Best-bases feature extraction algorithms for classification ofhyperspectral data: Geoscience and Remote Sensing, IEEE Transactions on, 39, 1368-1379.

J. J. Liu, D. Kim, and C. Han, 2007, Use of Wavelet Packet Transform in Characterization of Surface Quality: Ind. Eng. Chem. Res., 46, 5152-5158.

S. Livens, P. Scheunders, G. van de Wouwer, and D. Van Dyck, Year, Wavelets for texture analysis, an overview: Image Processing and Its Applications, 1997., Sixth International Conference on, 581-585 vol.582.

N. M. Rajpoot, Year, Texture classification using discriminant wavelet packet subbands, III-300-III-303 vol.303.

N. Saito, and R. R. Coifman, 1995, Local discriminant bases and their applications: Journal of Mathematical Imaging and Vision, 5, 337-358.

1497

Algorithm for Integrated Production and Utility System Scheduling of Batch Plants

Mujtaba H. Agha[a], Raphaele Thery[a], Gilles Hétreux[a], Alain Hait[b], Jean M. Le Lann[a]

[a]*Laboratoire de Génie Chimique (UMR 5503), 5 rue Paulin Talabot, 31106 Toulouse.*
[b]*Institut Supérieur de l'Aéronautique et de l'Espace, 10 av. E. Belin, 31055 Toulouse*

Abstract

Onsite utility systems feature in many industrial units. Due to technological advances in utility generation technologies, even energy intensive batch plants are attracted towards using onsite utility system. However in batch plants emphasis is placed solely on production process and utility aspects are not incorporated into the general scheduling model. This article presents an extended resource task network (ERTN) representation that can be used to develop a universal scheduling model that would concurrently undertake scheduling of production unit as well as an onsite utility system.

Keywords: Short-term scheduling, Utility system, Network representation, MILP.

1. Introduction

Increased competition and shrinking profit margins has forced chemical processing and manufacturing related industries to look for ways of improving productivity and reducing operational costs. To achieve these objectives, development of scheduling and production planning models has received increased attention. Scheduling can be defined as allocation of limited resources to produce one or more products. This allocation answers three primary questions – when (at what time), where (in which equipment) and how much (quantity) of material needs to be processed to produce desired products. Mathematical programming, especially Mixed Integer Linear Programming (MILP), because of its rigorousness, flexibility and extensive modeling capability is widely used for scheduling problems [Floudas & Lin, 2005].

Multiproduct / multipurpose batch plants are inherently flexible and their scheduling is dependent on the nature of product demand. When reliable forecasting can not be established then production is only driven by the available orders leading to short-term scheduling problem [Mendez *et al.*, 2006].

State Task Network (STN) proposed by Kondili *et al.* [1993] was a major step towards developing a universal short-term scheduling model for a batch plant. STN proposed a generalized framework for presenting the production process which was then used to develop a mathematical formulation applicable to all types of batch plants. Pantelides [1994] proposed Resource Task Network (RTN) which is extended framework for presenting the production process. RTN contains more information about processing equipment and their utilization, an aspect which is not explicitly shown in STN.

2. Problem Statement

Although STN and RTN are equally good at handling scheduling problems, but in both representations utility is simply considered as an external resource which allows

processing equipment to transform material from one form to another. The limited availability of utility is incorporated either as a scheduling resource constraint or as a term in objective function. This is not sufficient as unlike the resources classically considered in the scheduling problems (machinery or work force), utility has special characteristics which must be taken into account. Utility is present in various forms (electricity, steam at different pressure levels, hot/cold water, etc). It is also a resource which must be consumed immediately after its generation as generally it can not be stored in its ultimate useful form. Moreover, in case of an onsite utility system, the utility availability at a given time is dependent on the intrinsic functioning of the utility system. In consequence, the scheduling of the utility system is intertwined with the scheduling of the production process. Thery *et al.* [2008] demonstrated that significant gains in operational costs can be achieved by using an integrated production unit and utility system scheduling model. However, their formulation was too specific and lacking on the aspect of universality that is present in STN and RTN representations. Hence, there is a need to develop a broader universal framework that would incorporate the operational planning of utility system considerations into scheduling problem.

This work aims to attain this objective by developing an extended resource task network (ERTN) representation which can be used to carry out the scheduling of production unit and utility system simultaneously in a single universal scheduling model.

3. Extended Resource Task Network (ERTN) Representation

The main feature of ERTN is unified representation of production unit and onsite utility system. In ERTN various nodes and arcs are connected in such a manner that physical illustration of industrial process is attained. These nodes and arcs are described below:
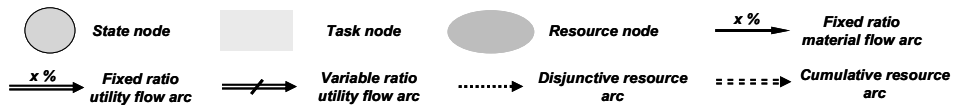
Figure 1: Nodes and arcs used in ERTN representation

3.1. Nodes used in ERTN

ERTN uses the same types of nodes as employed by its predecessor STN and RTN representations. The state node (circle) represents materials (feeds, intermediate and finished products) and the utilities (water, steam, electricity, fuel, etc). It is important to point out that the same state node can be considered as material resource (for onsite utility system) and as a utility resource (for the production system). The task node (rectangle) represents a process operation which transforms material or utility from input state(s) into output state(s). The resource node (oval) represents the processing equipment in which a particular processing operation, i.e., a task can be undertaken.

3.2. Arcs used in ERTN

Arcs are used to provide the necessary connection between the state, task and resource nodes of ERTN. Contrary to the RTN representation, an explicit distinction is made between matter flow (materials and utilities) and resource flow. Before discussing the respective arcs it is essential to outline the concept of principal and support utility that is used by the ERTN framework. The 'principal utilities' comprises of water and its derivatives (steam at various pressures, hot and cold water) while all other utilities (electricity, fossil fuel, etc) are considered as 'support utilities'.

In the ERTN framework the flow of material between task node and state node is represented by material arc. Similarly the flow of principal utility between task node

and state node is represented by utility arc. The material and principal utility flows can be either in fixed ratio of task batchsize or in variable ratio of task batchsize. The cumulative resource arc principally represents flow of different utilities which are generated by utility system and then consumed by production unit. On the other hand, the disjunctive resource arc just identifies the resources (in our case processing equipments) where a particular task can be undertaken.

3.3. Demonstration of ERTN representation through an example

To illustrate how ERTN can be constructed, this article uses example of an industrial unit comprising of a batch production unit and an onsite utility system (figure 2).
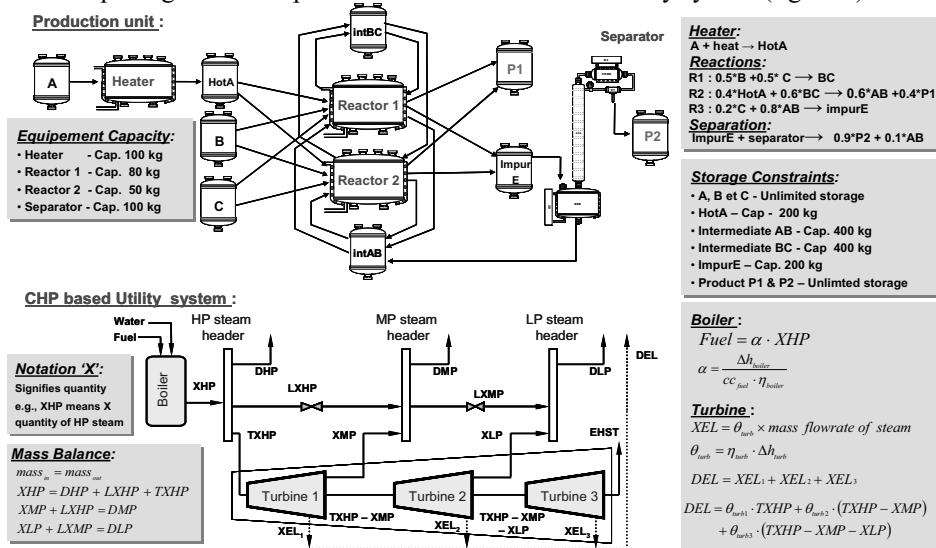
Figure 2: A batch production unit with a CHP based onsite utility system

The production unit converts three feeds (*A*, *B* & *C*) into four intermediate products (*HotA*, *intBC*, *intBC* & *ImpurE*) and two finished products (*P1* & *P2*). To achieve this, four process equipments (*Heater*, *Reactor 1*, *Reactor 2* & *Separator*) are used which consume following utilities – electricity (*DEL*), steam at high pressure (*DHP*), medium pressure (*DMP*) and low pressure (*DLP*). Three reactions (*R1*, *R2* & *R3*) can take place either in *Reactor 1* or in *Reactor 2*. To produce these utilities an onsite Combined Heat & Power (CHP) based utility system is used. The boiler consumes fuel to transform water into HP steam. HP steam is converted into MP and LP steam by using multistage turbine or by using pressure release valves. The advantage of using turbine is that it not only reduces pressure but also simultaneously generates electricity. The stages of turbines are represented as *turbine 1*, *turbine 2* & *turbine 3*.

The ERTN representation of above mentioned industrial unit is shown in figure 3. The upper half depicts the production unit while the lower half depicts utility system. The link between the two units is provided by flow of utilities *DEL, DHP, DMP* and *DLP*.

It is important to clarify the need of a cumulative resource arc. The utility arc simply illustrates the flow and transformation of principal utility from one phase to another within the utility system. However during this transformation process a task in utility system might not only consume support utility (e.g., task *T9* consuming *fuel* to convert water into *HP* steam) but it can also generate support utility (e.g., task *T11* not only converts *HP* steam into *MP* steam but it also generates *electricity*). As a result, the notion of cumulative resource arc is introduced which would incorporate the use of

support utilities without violating utility mass balance. This concept is further extended to include flow of utilities from utility system to production unit. The utility states providing utilities are considered as additional resources to the production unit and flow from them is represented as cumulative resource arc.
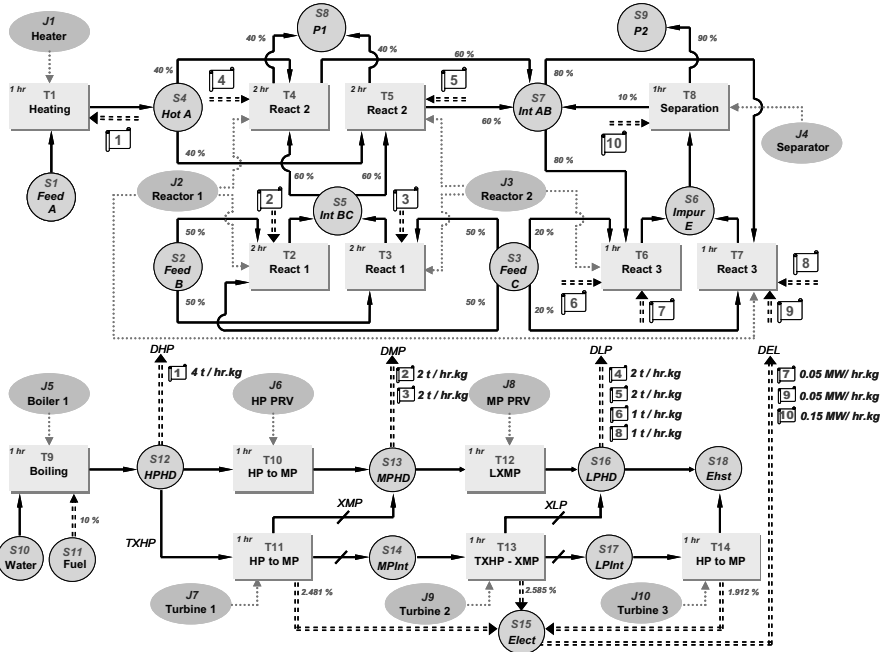
Figure 3: Extended resource task network (ERTN) representation of industrial unit

3.4. Mathematical formulation

To formulate the scheduling model, discrete time based Mixed Integer Linear Programming (MILP) is used. This model is established on following hypotheses:

1. Processing times for each task are known and are independent of batch size.
2. Once a task has been started it can not be stopped before its completion.
3. A disjunctive resource can be shared by several tasks. However, during the same time period only one task can be undertaken in a resource.
4. A task consumes one or more type of utilities dependent on the batch size undertaken.
5. A state can have multiple output and input streams. However, all the input streams must be of same quality. If different quality streams need to be mixed then this would constitute as a separate task.
6. A state can be considered as a storage station where material / utility can be stored. However, only primary energy (e.g., fuel) can be stored in utility state. Other utilities (steam, electricity, etc) can not be stored in their ultimate useful form.
7. The material inputs and outputs from a task in production units are fixed and known proportion of its batchsize. However, the utility inputs and outputs from a task in utility system are dependent on the utility demands.

$UVin_{i,s} \neq 0\ if$ ====▶ *enters into a task ; otherwise* = 0 $UVout_{i,s} \neq 0\ if$ ====▶ *leaves the task ; otherwise* = 0

$\rho_{i,s} = !0$ ——▶ *enters or leaves the state ; otherwise* = 0 $\alpha_{i,s} = 1, if$ ⇉ *enters or leaves the state ; otherwise* = 0

$\mu_{i,s} = 1, if$ ⇉ ====▶ *enters or leaves the state ; otherwise* = 0

Figure 4: Proportions signified by each arc

Equation (1) represents allocation constraints, that a resource *j* can only initiate task *i* at a given time *t*. It also illustrates that the resource *j* is not available to perform another task during the periods $t' = t - p_i + 1$ till $t' = t + p_i - 1$ (i.e., duration of the task). Equations (2) and (3) show production capacity and storage limitation constraints respectively. Equations (4) accounts for the material and utility mass balance across all the states in ERTN. Equations (5) and (6) represent the utility mass balance across the tasks in utility system. Equation (7) illustrates the utility demands from utility system by tasks in production unit. Equation (8) accounts for the support utilities generated by the tasks in utility system. Finally, Eq. (9) corresponds to objective function that is minimized to achieve the short-term schedule. The objective function is a weighted sum of electricity purchase cost and storage parameters for tardiness starting date.

$$\sum_{i \in I_j} \sum_{t'=t-p_i+1}^{t} W_{i,t} \leq 1 \qquad \forall j \in J, \ \forall t \in 1,..,T \tag{1}$$

$$W_{i,t} V_i^{\min} \leq B_{i,t} \leq W_{i,t} V_i^{\max} \qquad \forall i \in I, \ \forall t \in 1,..,T \tag{2}$$

$$0 \leq S_{s,t} \leq C_s^{\max} \qquad \forall s \in S, \forall t \in 1,..,T \tag{3}$$

$$S_{s,t} = S_{s,t-1} + \sum_{i \in I} \rho_{i,s}^{prod} B_{i,t-p_i} + \sum_{i \in I} \mu_{i,s}^{prod} BUout_{s,i,t} - \sum_{i \in I} \rho_{i,s}^{cons} B_{i,t} - \sum_{i \in I} \mu_{i,s}^{cons} BUin_{s,i,t}$$
$$+ imt_{s,t} - ext_{s,t} \qquad \forall s \in S, \forall t \in 1,..,T \tag{4}$$

$$B_{i,t} = \sum_{s \in S_{util}} \alpha_{i,s}^{prod} BUout_{s,i,t} \qquad \forall i \in I_{util}, \forall t \in 1,..,T \tag{5}$$

$$B_{i,t} = \sum_{s \in S_{util}} \alpha_{i,s}^{cons} \cdot BUin_{s,i,t} \qquad \forall i \in I_{util}, \forall t \in 1,..,T \tag{6}$$

$$BUin_{s,i,t} = uf_{i,s} W_{i,t} + UVin_{i,s} \sum_{t'=t-p_i+1}^{t} B_{i,t'} \qquad \forall s \in S_{util}, \forall i \in I, \forall t \in 1..T \tag{7}$$

$$BUout_{s,i,t} = uf_{i,s} W_{i,t} + UVout_{i,s} \sum_{t'=t-p_i+1}^{t} B_{i,t'} \qquad \forall s \in S_{util}, \forall i \in I, \forall t \in 1..T \tag{8}$$

$$Funobj = \sum_{t} \sum_{s \in S_{elec}} electricity\ price \cdot imp_{s,t} + \sum_{t} \sum_{s \in S} h_s \cdot ext_{s,t} + \sum_{t} \sum_{s \in S} \frac{h_s \cdot S_{s,t}}{1000} \tag{9}$$

3.5. Results and concluding remarks

To judge effectiveness of ERTN the above mentioned example was formulated and resolved using software XPRESS-MP. The problem was solved on an Intel(R) Core(TM) 2 Duo CPU @ 2.00 GHz and 1.00 GB of RAM. A 24 hour planning horizon is considered with the time intervals divided into 24 one hour periods. The demand of the finished product P1 & P2 is 300 and 250 kilograms respectively. The short-term schedule attained through ERTN representation is presented in figure 4. The schedule gives nature, time and batchsize of tasks undertaken in each resource of industrial unit. This article presents Extended Resource Task Network (ERTN) representation which can be used to solve the solving the short-term scheduling of the batch plant. The ERTN extends its predecessors network representation by incorporating the operational planning aspects of an onsite utility system. The use of ERTN representation results in a universal scheduling model which simultaneously performs scheduling of production unit and that of onsite utility system.

This study is a part of research being conducted at CNRS, whose objective is energy management solutions for mono and multi-sites. In future the discrete time model will be replaced by a continuous time formulation.
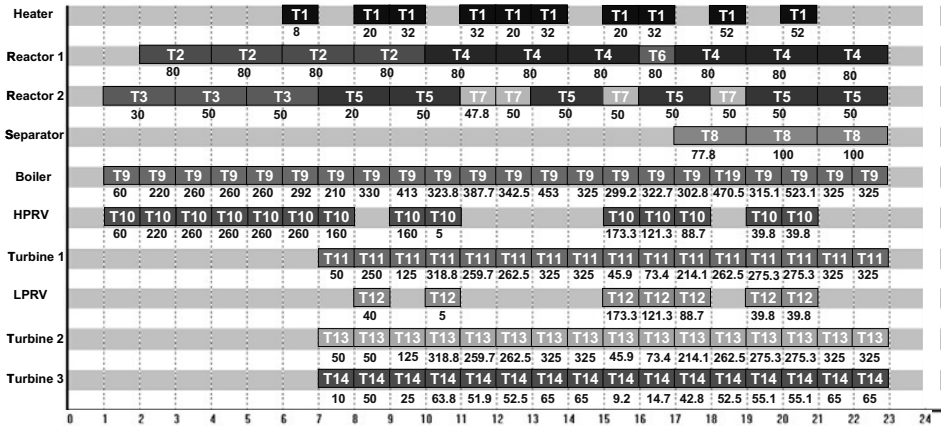
Figure 5: Scheduling of the industrial unit based on ERTN representation

Nomenclature

Indices i: tasks, s: states, j: resources, *boil*: boiler, *elec*: electricity, *mat*: material, *util*: utility

Sets I: Set of tasks, I_{util}: Set of tasks in utility unit, J: Set of resources, S: Set of states, S_{mat}: Set of material states, S_{util}: Set of utility states

Variables		Parameters	
$B_{i,t}$	Amount of material at time t being undertaken by task i	$\rho_{i,s}^{cons}, \rho_{i,s}^{prod}$	proportion of material of state s consumed or produced by task i
$imt_{s,t}$	Import into a state from external source	$\alpha_{i,s}^{cons}, \alpha_{i,s}^{prod}$	proportion of utility of state s consumed or produced by task i
$emt_{s,t}$	Export to a state from external source	$\mu_{i,s}^{cons}, \mu_{i,s}^{prod}$	proportion of utility directed into or from state s of utility system to task i in production unit
$BUin_{s,i,j,t}$	Utility input to a task	c_s^{max}	maximum storage capacity of state s
$BUout_{s,i,j,t}$	Utility output to a task	h_s	Coefficient for tardiness starting date
$S_{s,t}$	Storage of state s at time t	$uf_{i,s}$	Fixed utility demand of a task
$W_{i,t}$	Binary variable, with value '1' if task i is undertaken at time t, otherwise value is '0'.	$UVin_{i,s}$	Utility demand of a production task
		$UVout_{i,s}$	Support utility generated by utility task

References

C.A. Floudas & X. Lin, X, 2005. Mixed Integer Linear Programming in Process Scheduling: Modeling, Algorithms, and Applications, Annals of Operations Research, 139(1), 131-162.

E. Kondili, C.C Pantelides, W.H. Sargent, 1993, A general algorithm for short-term scheduling of batch operations–I. MILP formulation, Computers and Chemical Engineering, 17(2), 211-227.

C.A. Méndez, J. Cerda, I.E. Crossmann, I. Harjunkoski and M. Fahl, 2006, State-of-the-art review of optimization methods for short-term scheduling of batch processes, Computers and Chemical Engineering, 30, 913-946.

C.C. Pantelides, 1994, Unified frameworks for the optimal process planning and scheduling. In: CAChE Publications. Proceedings of the Second Conference on Foundations of Computer Aided Operations, 253-274.

R.Thery, G. Hétreux and M. Agha, 2008, Modèle intégré pour l'ordonnancement d'ateliers batch et la planification de centrales de cogénération. Proceedings of 7eme Conférence Internationale de Modélisation et Simulation (MOSIM 08).

1503

Operation and Composition Control of a New Pressure Swing Batch Distillation System

Arpad Kopasz, Gabor Modla, Peter Lang

BUTE Department of Building Services and Process Engineering, Muegyetem rkp. 3- 5. H-1521 Budapest,Hungary, E-mail: lang@mail.bme.hu

Abstract

The pressure swing separation of a binary minimum azeotrope (n-pentane-acetone) in a double column batch stripper is studied by rigorous simulation. For controlling the product compositions a simple scheme is presented. On the basis of temperatures of bottoms product PID controllers manipulate their flow rates varying the reboil ratios. The controllers are tuned. The influence of the most important operational parameter (division ratio of liquid leaving the common top vessel) is investigated. For rigorous simulation calculations a professional dynamic flowsheet simulator is applied.

Keywords: batch stripping, pressure swing, minimum azeotrope, control.

1. Introduction

Binary pressure sensitive azeotropes can be separated by pressure swing distillation (PSD). Continuous PSD was first applied in the industry in 1928. Phimister and Seider (2000) studied first the batch (stripping) and semicontinuous application of PSD by simulation. First Repke et al. (2007) investigated experimentally the batch PSD (PSBD, pilot-plant experiments for the separation of a minimum azeotrope in a batch rectifier (BR) and stripper (BS)). Modla and Lang (2008) studied different batch configurations (BR, BS, combination of BR and BS and middle vessel column(MVC)) by feasibility studies and rigorous simulation for the separation binary (max. and min.) homoazeotropes. By modifying the MVC, which has not been proven suitable for the PSBD, they suggested two new double column batch configurations: rectifier (DCBR, *Fig. 1a*) and stripper (DCBS, *Fig. 1b*). They compared the different configurations for a given set of operational parameters without optimisation and control. For min. azeotropes the best results (minimal specific energy consumption for the same quality products) were obtained with the DCBS and for maximum azeotropes with the DCBR, respectively. The columns of these configurations can be operated practically in steady state. Modla et al. (2009) studied the feasibility of batch PSD separation of most frequent types of ternary homoazeotropic mixtures.

When operating these new configurations the liquid composition of the common vessel of the two columns must be kept between the two azeotropic compositions. The ratio of two product flow rates of a DCBS can be changed by varying the /reboil ratios and/or the ratio of division of the liquid flow leaving the common vessel.

The goals of this paper are:

-to investigate the operation of the DCBS for the separation of a minimum azeotrope,

-to study a simple scheme for the control of product compositions (temperatures of bottoms product are controlled and their flow rates are manipulated) ,

-to investigate the effects and to determine the optimal value of the liquid division ratio.

The calculations were made for the mixture n-pentane-acetone by using a professional dynamic simulator (CCDCOLUMN).
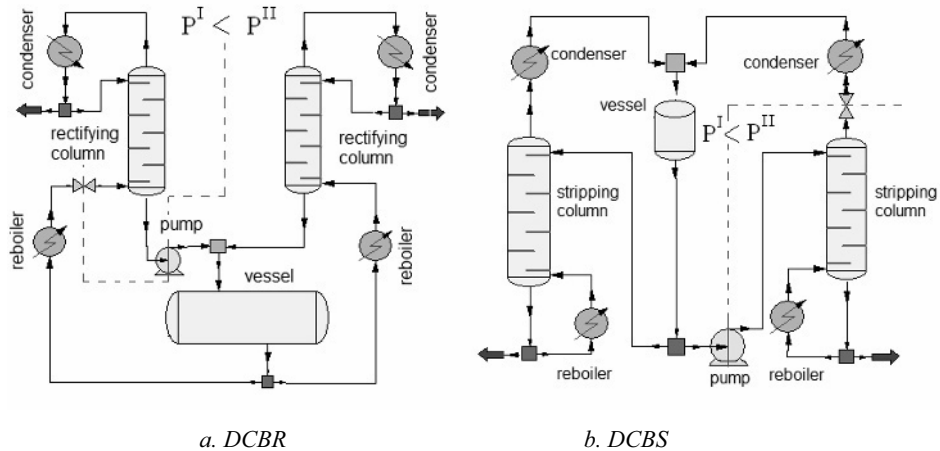
| | a. DCBR | b. DCBS |
Figure 1. The scheme of a DCBR and DCBS

The temperature-composition (*T-x,y*) diagrams and azeotropic data of the mixture studied are shown for the two different pressures in *Fig. 2* and *Table 1*, respectively.

Component		pentane(*A*)	acetone(*B*)
P	[bar]	1.01	10
x_{az}		0.754	0.668
T_{az}	[°C]	32.5	116.9
$T_{BP,A}$	[°C]	36.0	124.7
$T_{BP,B}$	[°C]	56.2	142.9
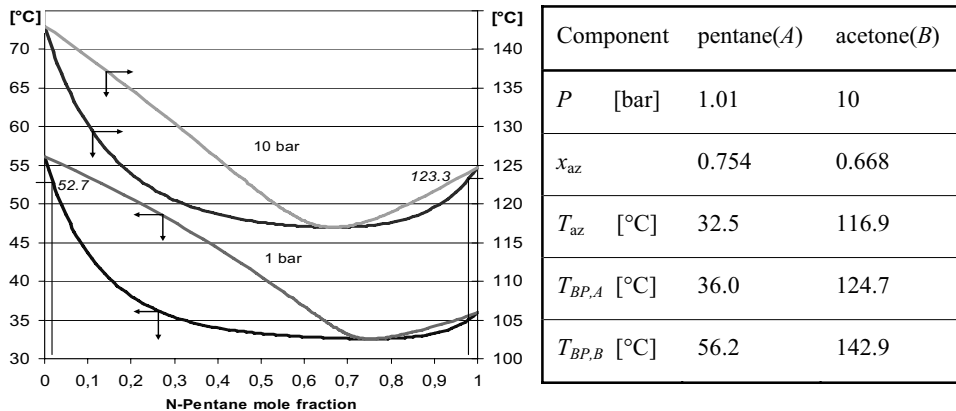
Figure 2.T-x,y diagrams of n-pentane-acetone

Table 1. Azeotropic data (UNIQUAC parameters: 571.98 and 95.033 cal/mol.)

2. Simulation method

The following simplifying assumptions were applied
- theoretical stages,
- negligible vapour hold-up,
- constant volumetric liquid plate hold-up.
The model equations to be solved are well known:
a. Non-linear differential equations (material balances, heat balances),
b. Algebraic equations (vapour-liquid equilibrium (VLE) relationships, summation equations, hold-up equivalence, physical property models).

For solving the above model equations we used the CCDCOLUMN dynamic flow-sheet simulator (ChemCad 6.0). For the simulation of columns simultaneous correction method was applied.

The following modules were used:

-DYNCOLUMN (column sections),

-DYNAMIC VESSEL (top vessel and product tanks),

-HEAT EXCHANGER, PUMP, VALVE,

-MIXER, DIVIDER,

-CONTROLLER, CONTROL VALVE.

The ChemCad model of the double column batch stripper with control of product compositions is shown in *Fig. 3*.
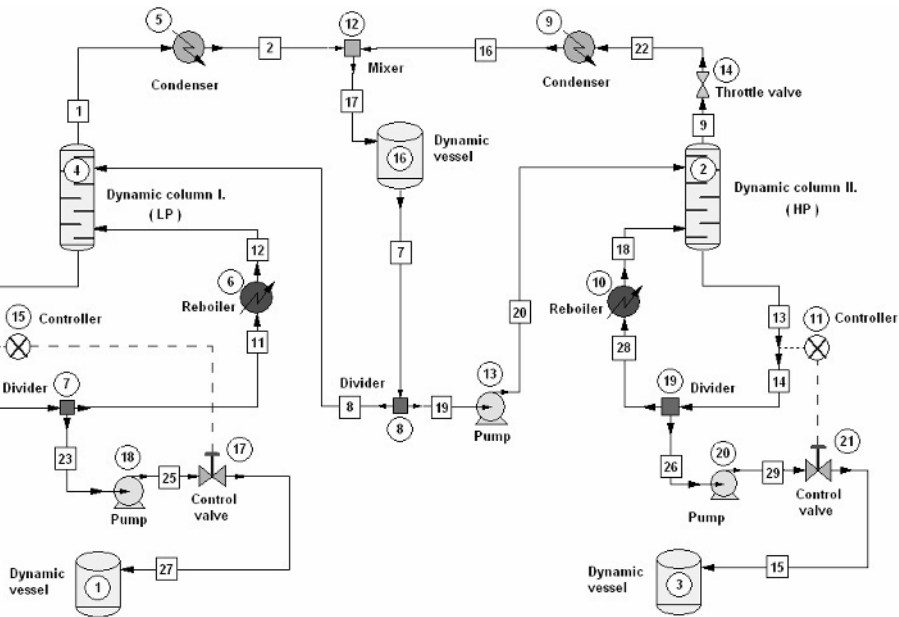
Figure 3. ChemCad model of the double column batch stripper with control loops.

3. Results

The number of theoretical stages for each column sections is 40. (The total condenser and total reboiler do not provide a theoretical stage.) The liquid hold-up is 2 dm3/plate. the pressure of the columns: $P^{LP}=1.013$ bar and $P^{HP}=10$ bar. At the start of the distillation plates of the columns are wet (they are filled with charge at its boiling point at the given pressure). The total flow rate of liquid leaving the common vessel: $L_{0,total} = L_0^{LP}+ L_0^{HP} = 6$ m³/h. The quantity of charge containing 30 mol% pentane is 4.022 kmol (0.471 m3). The prescribed purity is 98 mol% for both products. The reboil ratios R_s^{LP} and R_s^{HP} are changed by PID controllers manipulating (with linear control valves) the product flow rates W^{LP} and W^{HP}, respectively. The whole process is finished when the amount of liquid in the vessel decreases to 12.5 % of the charge.

First, the parameters of the two PID controllers (A_P, T_I and T_D) providing stable, good quality control of the product compositions in the whole region of liquid division ratio

($\phi=L_0^{LP}/L_{0,total}$) studied are determined. Then, the influence of this operational parameter on the performance of the PSBS is studied and its optimum value yielding the minimal overall specific energy consumption $((SQ^{LP}+SQ^{HP})/(SW^{LP}+SW^{HP}))$ is determined.

3.1. Tuning of PID controllers

Our aim is to determine a set of parameters of the PID controllers which provide good quality control of product compositions by taking into consideration the usual criterions (maximal overshoot, control time, number of oscillations).

The quality of control is determined by the evolution of not only the controlled variables (temperature of the two bottoms products) but also that of the position of the two control valves (varying the flow rate of the two bottoms product). The following criteria of quality of control are given concerning the two control valves:

-maximal overshoot: 33 %,

-maximum number of oscillations during the settling time T_S (within an error band of +- 5 %): 3.

In *Fig. 4* the evolution of the position of the control valve and bottoms composition of Column I for an inappropriate set of controller parameters is shown. (Both the position of the control valve and the controlled variable show oscillations without damping.)
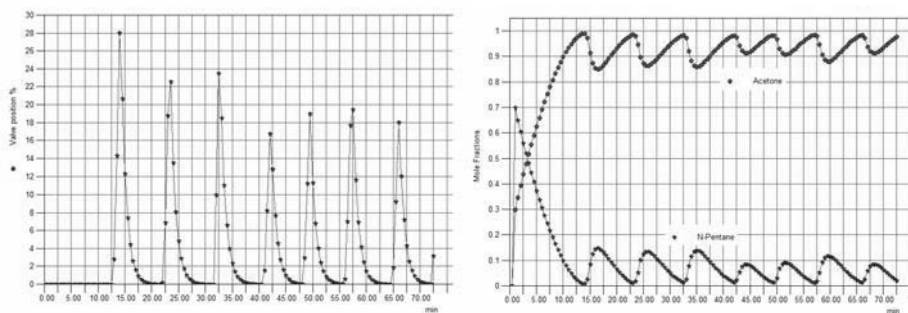
Figure 4. The evolution of the position of the control valve and bottoms composition for an inappropriate setting of PID parameters (Column I, PB=10 %, T_I= 1 min, T_D= 0)

Table 2. Parameters and quality data of control for an appropriate tuning

a. PID parameters:

	PB, %	T_I, min	T_D, min	Set point,°C
Column I	45	13	0.50	52.7
Column II	120	3	1	123.3

b. Valve flow coefficients: K_v^I= 0.13 K_v^{II}=2

c. Control quality data:

	Column I:	Column II:
Maximal overshoot:	(99.13-74.45)/74.45=**0.33**	(21.97-16.64)/)/16.64=**0.32**
Settling time, min:	17.25-10.5=**6.75**	17.75-10=**7.75**
No. of oscillations within T_S:	1	1

For the controller parameters selected (*Table 2*) the evolution of the position of the control valves, bottoms compositions and reboil ratios is shown in *Fig. 5*. (*Table 2* contains the control quality data, as well.)
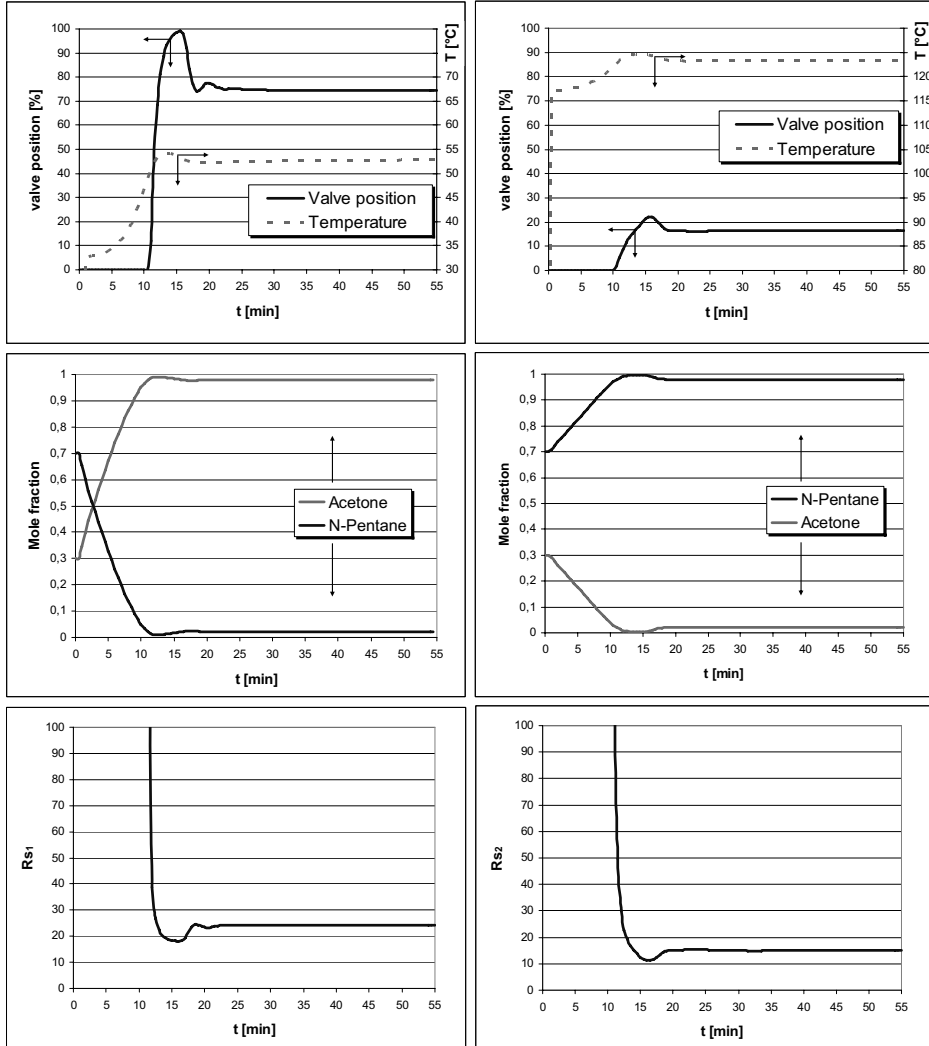
a. Column I   *b. Column II*

Fig. 5. The evolution of control valve positions and bottom temperatures (a), bottoms compositions (b), reboil ratios (c) (ϕ=0.55)

3.2. Influence of the liquid division ratio

The liquid division ratio is varied in the region 0.3-0.9. The specific energy consumption is minimal at ϕ=0.55 (*Fig. 6*). Prescribed purity products are obtained with reasonable recovery (*Table 3*). This table contains also the most important results for the process, such as the total and specific energy consumptions of the production. It must be

still noted that the recoveries could be still increased by reducing the quantity of residue in the common top vessel. In the case studied we were able to practically empty the vessel while maintaining the prescribed purities in the product tanks. However under a certain amount of residue (12 % of the charge) the operation of the control loops became unstable, therefore we stopped the process.
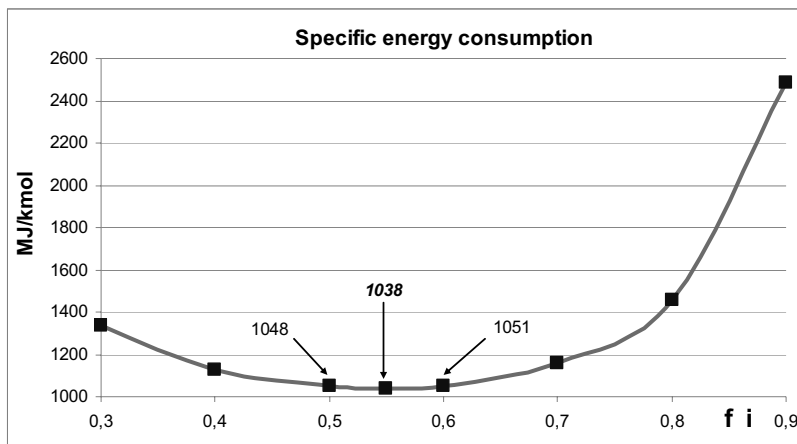
Fig. 6. The influence of the liquid division ratio on the specific energy consumption

Table 3. Most important results of the production for the optimal liquid division ratio

N-pentane recovery	%	75.32
Acetone recovery	%	67.54
N-pentane purity	mol %	98.20
Acetone purity	mol %	98.03
Total energy (SQ)	MJ	3106
Specific energy: $SQ/(SW_A+ SW_B)$	MJ/mol	1038
Production time	min	54

Acknowledgement

References

Modla G., A. Kopasz A. and P. Lang (2009). Feasibility of separation of ternary mixtures by pressure swing batch distillation, *PSE2009*.

Modla G. and Lang P. (2008). Feasibility of new pressure swing batch distillation methods, *Chem. Eng. Sci.*, 63, 2856-2874.

Phimister, J.R.; Seider, W.D. (2000). Semicontinuous pressure swing distillation, *Ind. Eng.Chem. Res.*, 39, 122-130.

Repke .J. U., Klein A., Bogle D., Wozny G.,(2007). Pressure Swing Batch Distillation for Homogenous Azeotropic Separation", *Chem. Eng. Res. Des.*, 85, 492-501.

1509

Monitoring Pipelines Through Acoustic Method.

Elisangela O. Sousa, Sandra L. Cruz, João A. F. R. Pereira

Department of Chemical Systems Engineering. School of Chemical Engineering
University of Campinas, UNICAMP. P.O. Box 6066, 13083-970, Campinas-SP, Brazil
E-mail address: eorlandi@desq.feq.unicamp.br

Abstract

Pipeline networks are complex systems of ducts used for liquid and gas transportation
through long distances. They frequently cross highly populated regions, water supplies
or natural reserves. Even small leaks in pipelines can lead to great losses of products
and serious damages to the environment before it could be detected. With the purpose to
track these leaks, a methodology is proposed for detection of leaks in pipelines based on
acoustic method and on analysis of pressure transients generated by leak occurrence.
Pressure transients and the sound noise generated by leakage are detected and analyzed
in a pipeline operating with continuous flow of gas (air) under various operation
conditions. The experimental results showed that it is possible to detect leaks in
pipelines based on hydraulic transients and on acoustic methods. The acoustic method
was decisive for leak detection; since the changes in pressure observed through the
pressure transducer was not significant depending on the gas flow rate. The analysis of
the signal amplitude for different frequencies shows that the leakage noise signal
changes with both leak magnitude and pipeline pressure
Keywords: leak detection, acoustic method, pipeline networks.

1. Introduction

Pipeline networks are frequently used for transportation and distribution of liquid an
gas products. Transport pipelines can cover large geographical areas and can be several
thousand of kilometers long. They operate at relatively high pressures. Compressors at
the beginning of the line provide the energy to move the gas through the pipeline and
compressor stations are required at a number of points along the line to maintain the
required pressure. In general, these pipelines are made of steel pipe and are buried
below ground surface. The individual sections of pipes are joined by welding, and the
pipe is externally coated to protect against corrosion. Pipe sizes can be as large as 1.5 m
in diameter.

Shimanskii et al (2005) based on high temperature microphones, which respond
rapidly to the appearance of the acoustic noise due to a leak. The main problem of the
subsequent investigations was to develop a method for detecting a small leak at an early
stage of its development. A leak detection method using microphones allows contact-
free monitoring, which makes it possible to decrease several fold the number of sensors
required, and the microphones can be placed in freely accessible zones around the
piping, which makes it easier to service the microphones themselves and the piping.

Jing-pin et al (2006) analyzed the propagation characteristics of guided waves in
acoustic leak location in pipelines. Time frequency methods were used in the analysis of
acoustic leak signals. The mode components of the acoustic leak signal were obtained
based on the relation of time-frequency distribution of acoustic leak signal and the

dispersion curves of guided waves. The research can provide a guideline for the mode selection in pipeline leak location, and help improve the accuracy of leak location.

Morozov et al (2007) described the first domestic multi-channel automated acoustic system for monitoring leaks, called SAKT. It was based on the generation of high frequency stress waves on the surface of pipes during outflow of a liquid which is under high pressure and high temperature through a rupture in the form of a crack.

Verde et al (2007) presented a method for the identification of two leaks in a pressurized single pipeline where both transient and static behavior of the fluid in leaks conditions are used to identify the parameters associated to the leaks without requirements of valve perturbation. They used an iron galvanized pipeline with 132.56 m long with a diameter of 0.105 m, thickness of 4.7 mm. Two sonic sensors had been installed extreme to measure flows of the static and transient state of the liquid. The measures of pressure had been carried through a pressure transmitter and the signals had been measured using a system of acquisition of data with a frequency of 100 hertz. The results had shown that to detect multiple leaks in conditions of static state of the liquid the test with a dynamic model is necessary complementary to reduce the time interval expense ploughs the identification of the leaks.

2. Experimental Work

The rigid pipeline consisted of a bunch of ½" in diameter, 60 m long galvanized iron tube. Figure 1 show the experimental assembly used to simulate the occurrence of leaks in the rigid pipeline.
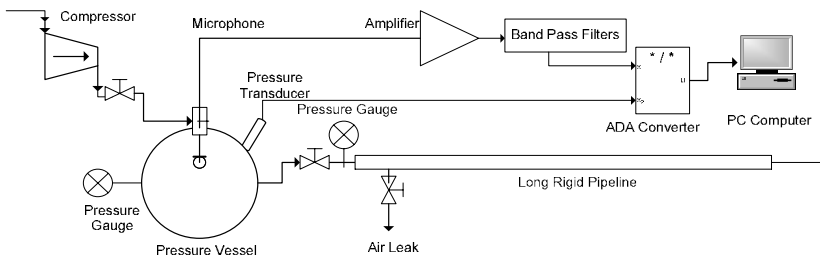
Figure 1. Rigid gas pipeline experimental assembly

The flexible pipeline consisted of a rolled up 50 m long transparent rubber helical wired. Figure 2 show the experimental assembly used to simulate the occurrence of leaks in the flexible pipeline.
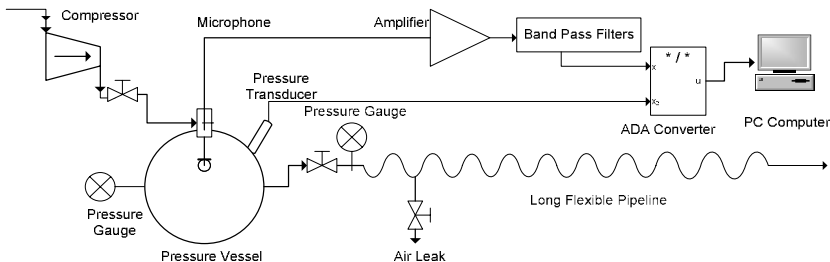
Figure 2. Flexible gas pipeline experimental assembly

Leaks were provoked manually through a rapid aperture valve installed in a side outlet at the entrance of the pipeline. The leak magnitude was controlled through orifice installed in the leak line. The orifice size varied from 0.1 mm to 5 mm in diameter.

The gas (air) used in the experiments was provided by the compressor which feed the air line of faculty laboratories and was fed to the pipeline through a 37.5 liters pressure vessel. Compressed air pressure was measured by manometers installed in the pressure vessel and in the pipeline. The air pressure (P0) varied from 1.0 to 7.0 kgf/cm^2.

In the leakage experiments with steady state gas flow a 0.8 mm orifice was installed at the end of the pipeline to keep it pressurized.

Pipeline monitoring was made through a microphone and also a pressure transducer installed in the pipeline, both connected to a PC computer through an ADA converter.

The microphone was developed in the laboratory to catch the sound noise generated by leak occurrence. On the first stage the signal was amplified and on the second one, the signal was filtered through three band pass filters, centered in 1 kHz, 5 kHz and 9 kHz each one, generating three continuous signals in different frequencies. A piezoelectrical transducer. The transducer was connected to the computer through an ADA converter.

The PC computer used for the data acquisition was a Pentium 233 MHz, 500MB and 16 MB RAM. The data acquisition software, written in C language, was developed to read and filter all the pressure transducers signals and to display the pressure transient and sound noise amplitude profiles plots.

3. Results and Analysis

3.1. Leak Detection in Rigid Pipeline

Figures 3 to 6 show the pressure transient and sound noise amplitude profiles caused by gas leak through orifices of 1.0 mm and 3.0 mm, when the rigid pipeline operated under constant pressure of 4.0 kgf/cm² and 6.0 kgf/cm².

When the leak occurred the pressure dropped, suddenly as the leak magnitude (orifice size) increased, while the sound noise amplitude increased rapidly, both remaining practically constant during leak occurrence. In Figure 1 the sound amplitude was about 1 V and 2 V in Figure 2. With the orifice of 3.0 mm the sound amplitude was about 3.5 V when the pressure was 4.0 kgf/cm² and reached 4.75 V when the pressure was 6.0 kgf/cm².

The dominant frequency was 5 kHz, independently of the orifice size, indicating that leak occurrence generated a medium sound noise. The second frequency was a high frequency, 9 kHz, with the 1.0 mm orifice and low frequency, 1 kHz, with the 3.0 mm orifice, indicating a more serious sound.

The obtained results also show that leaks were readily detected through the sound noise amplitude changes, but that not always happened through pressure transient profiles.

The initial work pressure has no influence on the leak; the pressure drop is almost constant. This is considered as choked flow where de fluid at a certain pressure and temperature through a restriction, such as the hole, or orifice, or a valve in pipe into a lower pressure, limiting the air speed.
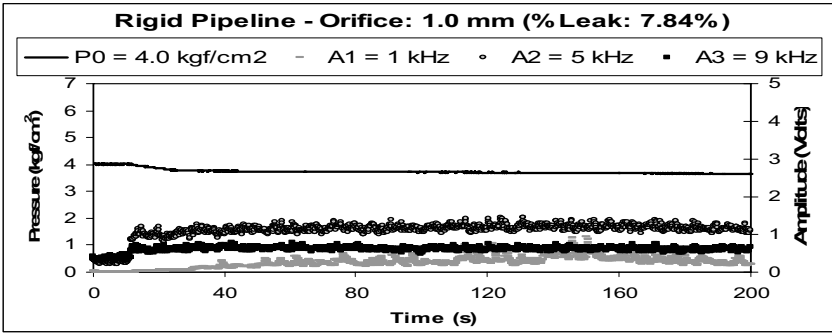
Figure 3. Pressure transients and sound amplitude. (Orifice 1.0mm). P0 = 4 kgf/cm^2.
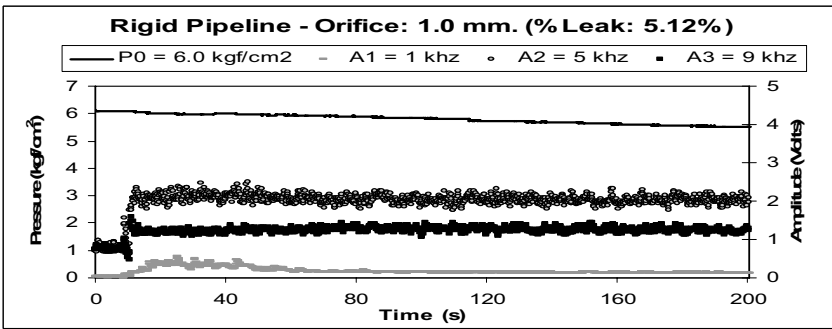
Figure 4. Pressure transients and sound amplitude. (Orifice 1.0mm). P0 = 6 kgf/cm^2.
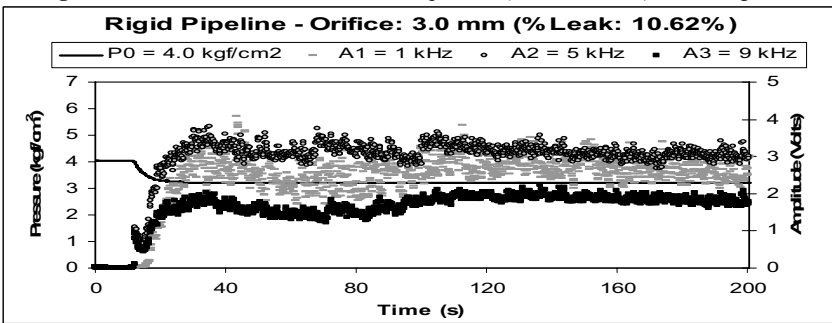
Figure 5. Pressure transients and sound amplitude. (orifice 3.0mm). P0 = 4.0 kgf/cm^2.
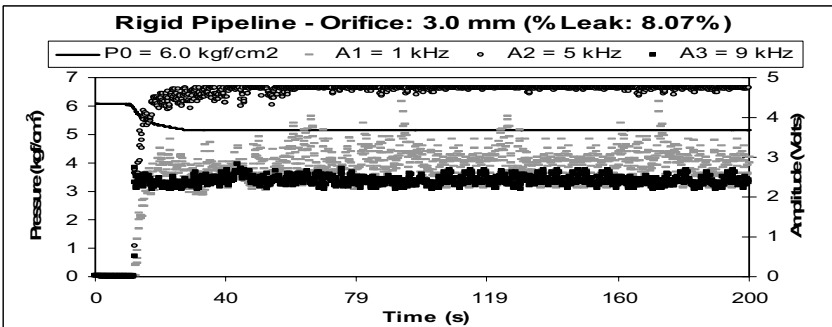
Figure 6. Pressure transients and sound amplitude. (orifice 3.0mm). P0 = 6.0 kgf/cm^2.

3.2. Leak Detection in Flexible Pipeline

Figures 7 to 10 show the pressure transient and sound noise amplitude when leak occurred through orifices of 1,0 mm and 3.0 mm and the flexible pipeline operated under constant pressure of 4.0 kgf/cm² and 6.0 kgf/cm².

The obtained results show that the pressure transient profiles presented the same behavior as that in the rigid pipeline. From the point of view leak detection it was not also be possible to detect the leak by using the pressure transient data for the 1.0 mm orifice.

The leak percentage calculated for the experiments showed that the leak percentage in flexible pipeline was lower when compared with the rigid pipeline, proving that when the leak is provoked in flexible pipeline, the pressure speeds of the wave has a damping. The graphics show that the sound noise amplitude was a few increasing and was not always a sudden change. The characteristics of sound noises were the same as in the rigid pipeline.

Figures 7 to 10 show that sound noise amplitude increase with both pressure and leak magnitude. With the orifice of 1.0 mm in diameter the sound amplitude was about 1 V when the pressure in the pipeline was 4.0 kgf/cm² and 1.5 V when the pressure was 6.0 kgf/cm². With the orifice of 3.0 mm the sound amplitude reached 2.5 V when the pressure was 4.0 kgf/cm² and reached 3.8 V when the pressure was 6.0 kgf/cm². These results show that the sound wave was attenuated in flexible pipeline.
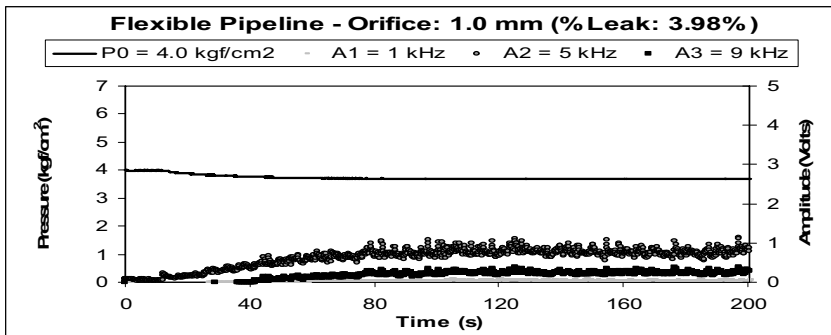
Figure 7. Pressure transients and sound amplitude. (orifice 1.0mm). P0 = 4.0 kgf/cm^2
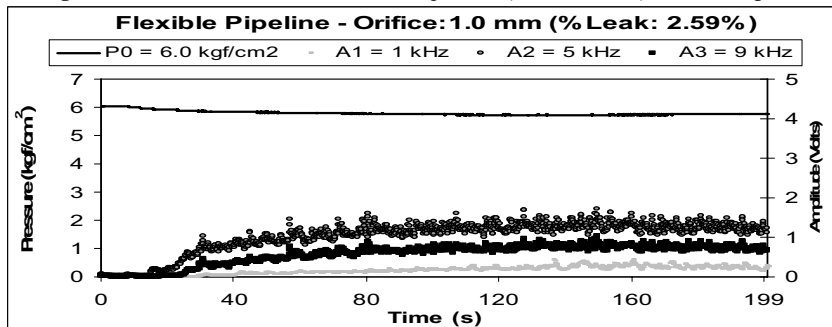
Figure 8. Pressure transients and sound amplitude. (orifice 1.0mm). P0 = 6.0 kgf/cm^2
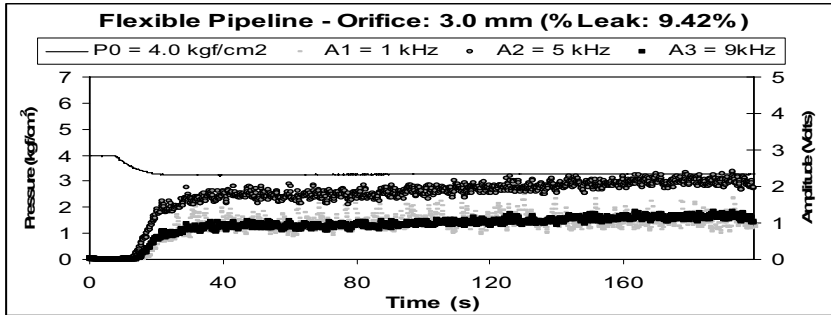
Figure 9. Pressure transients and sound amplitude. (orifice 3.0mm). P0 = 4.0 kgf/cm^2
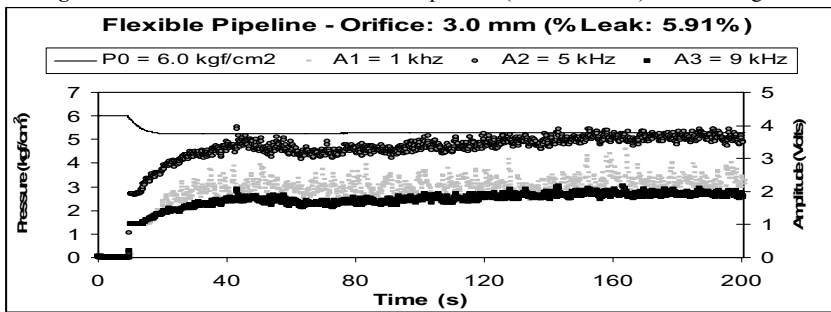
Figure 10. Pressure transients and sound amplitude. (orifice 3.0mm). P0 = 6.0 kgf/cm^2

4. Conclusions

A leak detection technique based on the sound noise generated by leak occurrence has been developed and tested to detect gas leaks in a long pipeline. The results showed that it is possible to detect leaks in pipelines based on acoustic method. In leak detection the presence of the microphone was decisive, since the changes in pressure measured through the pressure transducer may be negligible depending on the gas flow rate.

The analysis of the sound amplitude showed that the characteristics of leak sound noise changes with both leak magnitude and pipeline pressure. The present work is being continued so that the developed technique could become a tool for pipeline supervision.

References

Da Silva, H.V. Morooka, C.K. Guilherme, I.R. Da Fonseca, T.C. Mendes, J.R.P. Leak detection in petroleum pipelines using fuzzy system. *J Petrol Sci. Eng*. 49:223-38, 2005.

Morozov, S.A., Kovtun, S.N., Budarin, A.A., Dvornikov, P.A., Kurdryaev, A.A., Kondratovich, F.V., Shutov, P.S., Shvetsov, D.M., Konoplev, N.P.. Development of Acoustic Leak Monitoring System. *Atomic Energy*, vol. 103, no 6: 925 - 931, 2007.

Jing-Pin, J., Ren-Yuan, F., Cun-Fu, H. Bin, W. Modal analysis of acoustic leak signal in pipelines using time-frequency analysis. *Front. Mechanical engineering.* China 2, 146-150, 2006.

Shimanskii, S.B., Strelkov, B.P., Anan´Ev, A.N., Lyubishkin, A.M., Iijima, T., Mochizuki, H., Kasai, Y., Yokota, K., Kanazawa, J. Acoustic method of leak detection using high-temperature microphone. *Atomic Energy*. vol. 98, no 2: 89-96, 2005.

Verde, C. Visairo, N. Gentil, S. Two leaks isolation in a pipeline by transient response. *Advances in Water Resources*. 2007.

Timed Automata Models for Batch Scheduling with Sequence-Dependent Changeovers

Subanatarajan Subbiah[a], Sebastian Engell[a]

[a]*Process Dynamics and Operations Group,*
Department of Biochemical and Chemical Engineering,
Technische Universität Dortmund, Emil-Figge Straße 70, 44227 Dortmund, Germany.

Abstract

The standard approach to solve scheduling problems in the process industries is to use mathematical model formulations such as MI(N)LP. Recently, an alternative approach that has gained attention is to use reachability analysis for timed automata (TA) to solve such problems. In this contribution, we discuss an application of the TA based framework to model and solve batch scheduling problems with sequence-dependent changeovers. The resources and the jobs are modeled as sets of interacting TA in a modular fashion and are composed to form a global automaton which represents the complete model of the scheduling problem. The solutions for the scheduling problem are computed by performing a cost-optimal reachability analysis in the global automaton.

Keywords: Multi-product batch plants, sequence-dependent changeovers, scheduling.

1. Introduction

In multi-product and multi-purpose batch processing industries, where a variety of products have to be produced with scarce available resources, the problem of scheduling with sequence-dependent changeovers frequently arises. The processing units require changeover times to switch the production from one product to another. In the presence of significant sequence-dependent changeovers, the utilization times of the processing units are strongly influenced by the sequence in which the products are produced. In this case it is necessary to have efficient scheduling models that include setup times and sequence-dependent changeover times. Most of the solution approaches proposed in the last years solve such problems by modelling them as mathematical programming formulations (MILP or MINLP) and applying commercial solvers to solve them. The representation of time in the models plays a crucial role for the performance of the solvers. In [1], the authors propose a MILP formulation in discrete-time where the key feature is to use specific variables for each unit to track the unit-operation events. Continuous time formulations based on single-time grid, multiple-time grid, and slot based approaches can be seen in [2], [3] and [4]. In [5] the authors present a rule-based heuristic using a genetic algorithm to solve single-stage multiproduct batch processes with jobs up to 200 orders. For exact approaches such as MI(N)LP and CP the application is hindered by the effort needed to formulate mathematical models and requires experience in algebraic modelling.

An alternative approach to model scheduling problems is to use the framework of timed automata and to solve the optimization problem using reachability analysis. Timed Automata (TA) are finite state automata extended by the notion of clocks to model discrete event systems with timed behavior. The formalism of TA has been

originally proposed in [6] and has been extended with the notion of costs, referred to as priced or weighted timed automata in [7]. Previous work on the TA based approach to scheduling problems with makespan minimization on hard job shop benchmarks were reported in [8] and [9]. A particular appeal of this approach is the modular and partly graphical modeling which enables inexperienced users to build models. Another advantage is the availability of powerful search algorithms that can be modified and extended for special purposes.

1.1. Background of Timed Automata

A short and informal definition of timed automata is given here, for complete definition of the syntax and semantics please refer to [7]. A timed automaton is defined by a tuple. $A = (\ L\ ,\ C,\ \Theta,\ inv,\ l_0,\ F\ )$ in which:

- L represents the finite set of discrete locations, with l_0, $F \in L$, where l_0 represents the initial location and F represents the set of final locations.
- C represents the set of clocks assigned to the TA.
- $\Theta \subset L \times \gamma \times Act \times U(C) \times L$ represents the set of transitions between the locations where, γ is a set of guards specified as conjunctions of constraints of the form $c_i \otimes n$ or $c_i - c_j \otimes n$, where $c_i, c_j \in C$, and $\otimes \in \{\leq, ==, \geq, <, >, \neq\}$ and $n \in N$.
- Act represents the set of actions (e.g. invoking a new event or changing the value of a variable) while a transition is fired.
- $U(C)$ represents the set of clocks that are reset to zero after an enabled transition fires. A transition between a source location l and target location l' with a guard $g \in \gamma(C)$, performing an action $a \in Act$ and resetting the clocks $r \in U(C)$ is denoted by $(l,\ g,\ a,\ r,\ l')$. A transition can occur only when the guard conditions are satisfied and the invariant conditions of the target location evaluate to true.
- inv represents a set of invariants that assign conditions for staying in locations. The invariant conditions must evaluate to true for the corresponding location to be active. The automaton is forced to leave the location when the invariant evaluates to false.

In this contribution, we discuss the application of the TA framework to model and to solve batch scheduling problems with sequence-dependent changeovers. Modelling a scheduling problem using the TA framework is performed in a modular fashion. The resources and the jobs are modelled individually as sets of interacting TA. The interactions between the sets of automata are established by synchronizing the transitions between the automata. Synchronized transitions have the same synchronization labels and fire only simultaneously in the corresponding automata. The sets of interacting automata are composed using the procedure named *parallel composition* thereby forming one composed global automaton. Usually, the parallel composition is only performed on-the-fly to reduce space complexity of the reachability analysis. The composed automaton represents the complete model of the problem and a cost-optimal reachability analysis is performed to derive the schedules. The reachability analysis starts from the initial location of the composed automaton that represents the initial state of the system and evaluates the successor states created by a successor relation. This enumerative process is continued until the specified final target location is reached with minimal cost. In the search, various reduction techniques can be used in order to explore the state space quickly [8, 9].

2. Case study

The TA based framework explained above is applied to the medium size ice-cream manufacturing plant described in [10]. The case study is a two-stage process where 8 different end-products (*A - H*) have to be manufactured in batches. A schematic representation of the process is shown in Fig. 1. In the first stage, the raw materials stored in the warehouses are transported to the mixing department where they are processed according to the product recipes and stored in intermediate storage units. The intermediate products are cooled down and after a minimum waiting period has elapsed, the frozen products are transferred to the packing lines. In the second stage the products are packed and delivered to the customers. The intermediate products are unstable and have a maximum shelf time period within which they should be transferred to the packing lines from the storage units.

The plant consists of a single processing line with a production rate of 4500kg/Hr and 6 intermediate storage vessels (S_1 - S_6) in the first stage. The vessels S_1 and S_2 have maximum capacity of 8000kg each and the vessels S_3 - S_6 have maximum capacities of 4000kg each. The second stage consists of 2 packing lines P_1 and P_2 with production rate depending on the product packed. A routing constraint exists such that the two 8000kg vessels are coupled only to packing line P_1 and the four 4000kg vessels are coupled only to packing line P_2. The products (*A -D*) are packed only in packing line P_1 and products (*E -H*) are packed only in packing line P_2. Thus products (*A - D*) can be stored in storage vessels S_1 and S_2 only and products (*E - H*) can be stored in storage vessels S_3 - S_6 due to the routing constraint stated above. For the rest of the paper we consider that one batch of product measures 4000kg and the task durations are considered in minutes. Different instances of the case study are considered where production orders with different amounts of various products have to be produced within the minimum makespan, where a production order is equal to one batch of the corresponding product. The main features of the case study are: (a) the processing units in the first stage and the packing lines in the second stage are subject to sequence-dependent setup times and changeover procedure, (b) the intermediate storage units are limited have to be shared and (c) the intermediate products are unstable and have a minimum waiting time and a maximum shelf-life period, introducing a timing constraint between the processing and the packing task.
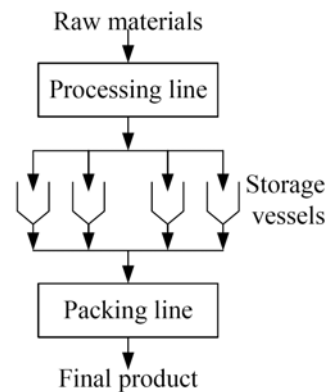
Raw materials

Processing line

Storage vessels

Packing line

Final product

Figure 1. Process flow diagram

3. TA model for the case study

This section explains the framework to derive a TA model for the scheduling problem by constructing interacting TA for the recipes and for the resources.

3.1. Recipe automata

For each production order defined by its recipe number, one *recipe automaton* and a clock c_i are used. The automaton for recipe *A* is shown in Fig. 2. Each operation is defined by a *waiting* location and by an *executing* location connected by transitions. Starting the execution of the mixing task by occupying the processing line is represented by the transition α_{1A} between the locations *Wait Proc_A* and *Exec Proc_A*. Finishing the execution of the task in the processing line and transferring the intermediate material to one of the storage units S_1 or S_2 is modeled by the transition
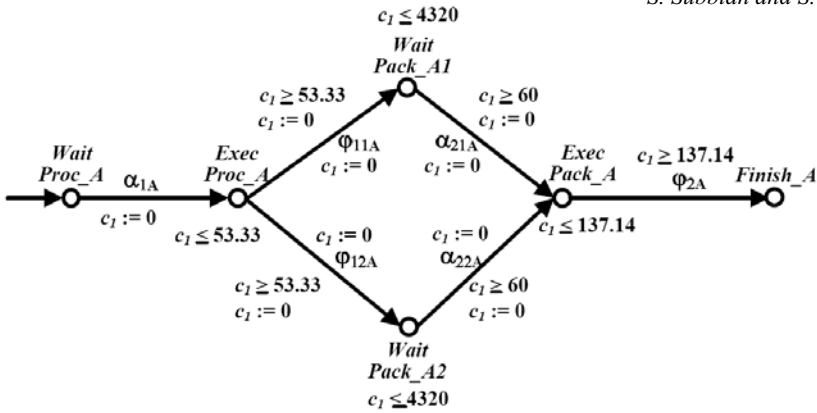
Figure 2. Recipe automaton for product A

φ_{11A} and φ_{12A}, respectively. Locations *Wait Pack_A1* and *Wait Pack_A2* represent that the intermediate material is stored in the storage unit S_1 and S_2, respectively and waiting to be packed in the packing line P_1. Starting the execution of the packing task in P_1 by consuming the intermediate material from the storage unit S_1 and S_2 is represented by α_{21A}, and α_{22A}, respectively. The additional location *Finish_A* represents the completion of the production order to produce one batch according to recipe A. The clock c_1 is introduced to model the timing behavior of the recipe. The invariants in the execute locations of the tasks force the automaton to leave the location once the task durations have expired. The guard conditions on the transitions that represents finishing of tasks which are labeled with φ ensures that the task is executed for the corresponding duration only. The clocks are reset to zero at every transition. The recipes A, B, C and D have the same automaton structure as the recipe structure is identical. Similarly the recipes E, F, G and H share the same automaton structure.

3.2. Resource automata

For each resource, a separate resource automaton with a clock is created by defining locations *idle* and *busy* for all possible configurations and operations which the resource executes. The automaton for the processing line with tasks representing processing of recipes A, B and C is shown in Fig. 3(a). The changeover table shown in Fig. 3(b) defines the changeover times (e. g. a minimum time period of CO_{AB} is required to changeover from configuration A to configuration B).

To From	idle	Product A	Product B	Product C
idle	-	Set_A	Set_B	Set_C
Product A	$Clean_A$	-	CO_{AB}	CO_{AC}
Product B	$Clean_B$	CO_{BA}	-	CO_{BC}
Product C	$Clean_C$	CO_{CA}	CO_{CB}	-

(a)   (b)
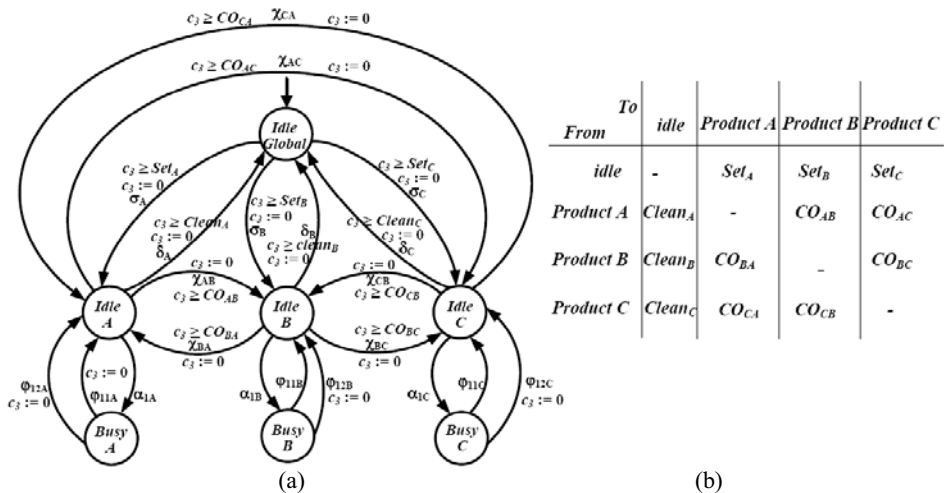
Figure 3. Resource automaton for the processing line and the changeover matrix.

     Transitions that represent the setting up of the resource from the global configuration to a particular configuration are labeled by σ and transitions that represent resetting or cleaning of the resource from a configuration back to the global configuration are labeled by δ. Transitions that represent changeovers between any two non-identical configurations are labeled by χ. The resource automata and the recipe automata interact by synchronized transitions using the synchronization labels α and φ. In the resource automaton, the α labeled transitions represent allocation of the resource and the φ labeled transitions represent the release of the resource. Guard conditions based on *clock* valuations on the setup transitions ensures that the transition can take place only after the respective defined setup time. Similarly, guard conditions on the cleanup transitions and changeover transitions ensure that the corresponding transition is enabled only after a minimal reset time or a minimal changeover time.

3.3. Storage automata

For each common intermediate storage vessel, a *storage automaton* is created with a shared integer variable V. The automaton for the storage vessel S_1 which is used as a common intermediate storage unit in the recipes A, B, C and D is shown in Fig. 4. The integer variable V_1 represents the number of batches of the material present in S_1 and b_i represents the amount of product i processed. The maximum capacity of the storage unit is depicted as (V_1). The *storage automaton* consists of an *Empty* location and, for each possible material that can be stored in the storage vessel, a *Filled* location is defined.
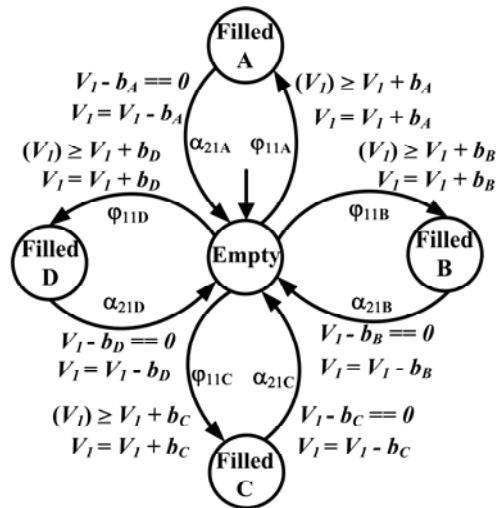
Figure 4. Storage automaton for the vessel S_1

     The *storage automata* interact with the resource automata and recipe automata using the synchronization labels and φ, where φ represents allocation of the storage unit and α represents release of the storage unit.

4. Computational experiments and results

The plant description with information on the resources, separate recipe files for 8 different recipes in the form of RTN and a table of production orders with reference to recipes are given as input to TAOpt, a tool developed at the Process Dynamics and Operations group [9]. From these inputs, TAOpt creates sets of recipe automata and resource automata in a modular fashion and the reachability analysis is performed on-the-fly. The search algorithm used to explore the reachability tree is a combination of depth-first search and best-first search. Various reduction techniques such as weak non-laziness and sleep-set method were used to reduce the search-space. A detailed description of the reachability analysis and of the search-space reduction techniques can be found in [9]. In addition to the above mentioned reduction techniques, based on assumptions on the setup times, cleanup times and changeover times for a resource with possible configurations i, j and k, partial traces in the reachability tree that lead to sub-optimal schedules are pruned. The assumptions on the setup times, changeover times and cleanup times for configurations i, j and k are as follows:

Table 1. Results for the test on makespan minimization: (T_{cpu}) computation time in CPU sec., (C_{max}.) makespan in minutes, (Created nodes) total number of nodes created to reach the solution, (Visted nodes) total number of nodes explored to reach the solution.

Orders	First feasible solution				Best feasible solution			
	T_{CPU}	Nodes		C_{max}	T_{CPU}	Nodes		C_{max}
		Visited	Created			Visited	Created	
10	0.98	37255	68240	1667.61	1.01	37984	69676	1607.21
20	2.24	73706	138425	2452.84	2.26	74382	139765	2392.84
32	5.23	109095	225957	4333.75	8.63	109154	226023	4273.72
46	6.32	112264	208412	5506.13	44.73	545750	1403457	5446.13
48	6.40	118452	216790	5780.41	24.63	405391	979564	5720.41

- $Setup_i + Changeover_{ij} > Setup_j$
- $Changeover_{ik} + Cleanup_k + Setup_j > Changeover_{ij}$
- $Cleanup_i + Setup_j > Changeover_{ij}$
- $Changeover_{ij} + Changeover_{jk} > Changeover_{ik}$

The TA based modeling and solution approach was applied to various problem instances of the case-study ranging from 10 to 48 orders. The computational equipment used for the tests is a 3.06 GHz Xeon machine with 2 GB memory. For all tests the computation time was restricted to 3000 CPU seconds and the node limit was restricted to 9 million to avoid memory overflow. The results obtained for various instances are shown in Table 1. For all instances a feasible schedule could be obtained within 7 CPU seconds. This shows the TA based approach can be used in online and reactive scheduling where it is important to derive good feasible schedules within short computation times. The node limit of 9 million caused the termination in all experiments. We observed that the search scheme initially explored sub-optimal regions of the solution space. We therefore plan to use rule-based strategies in the search algorithm to guide the exploration of the solution space to promising regions.

5. Summary and outlook

This work demonstrates an application of the TA based approach to model and to solve the problem of scheduling batch processes with resources subject to sequence-dependent changeovers. Various other objective functions, such as minimization of changeover costs, can also be accommodated in the proposed approach by extending the automata with the notion of costs using Priced TA. Future work will investigate problems with various other objective functions and to embed rule-based techniques into the reachability analysis of TA.

References

1. Jeffrey D. Kelly and Danielle Zyngier, 2007, *Ind. Eng. Chem. Res.*, 4964 - 4973.
2. Chi-Wai Hui, et al. , 2000, *Comp. Chem. Engg.*, 2705 - 2717.
3. Pedro M. Castro, et al. , 2006, *Ind. Eng. Chem. Res.*, 6210 - 6226.
4. M. Erdirik-Dogan and I. E. Grossmann, 2008, *Ind. Eng. Chem. Res.*, 1159 - 1183.
5. Yaohua He, Chi-Wai Hui, 2008, *Comp. Chem. Engg.*, 3067 - 3083.
6. R. Alur and D.L. Dill, 1994, *Theoretical Computer Science*, vol. 126, 183-235.
7. K. Larsen et al. , 2001, *Proc. 13th Int. Conf. on Computer Aided Verification,* 493 - 505.
8. Y. Abdeddaim and O. Maler, 2006*, Theoretical Computer Science,* Vol. 354, 272 -300*.
9. S. Panek, S. Engell, and O. Stursberg, 2006, *Control Engg. Practice*, 1183 - 1197.
10. Peter M. M. Bongers et al. , 2006*, Proc. 16th ESCAPE and 9th PSE,*1917-1922.

1521

Use of Event Correlation Analysis to Reduce Number of Alarms

Fumitaka Higuchi,[ac] Ichizo Yamamoto,[a] Tsutomu Takai,[b] Masaru Noda,[c]
Hirokazu Nishitani [c]

[a]*Idemitsu Kosan Co., Ltd, 2-1, Anesakikaigan, Ichihara, Chiba 299-0194, Japan*
[b]*Yamatake Corporation, 1-12-2 Kawana, Fujisawa 251-8522, Japan*
[c]*Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan*

Abstract

Event correlation analysis was applied alarm data at an Idemitsu Kosan ethylene plant in Japan to reduce the number of plant alarms. This is a data-mining method that detects statistical similarities among discrete occurrences of alarms or operations. By grouping correlated events based on degree of similarity, a policy for reducing alarms can be designed more easily than by analyzing individual alarms and operations. By using event correlation analysis, we were able to divide event data of alarms or operations of the ethylene plant into a limited number of groups. The results of the analysis were helpful for identifying unnecessary alarms, such as sequential alarms, buried in a lot of noisy plant data. This method is useful for reducing the number of alarms.

Keywords: Alarm Reduction, Top Ten Approach, Event Correlation Analysis, Ethylene Plant, EEMUA

1. Introduction

With the advance of distributed control systems (DCS) in the chemical industries, it has become possible to install many alarms cheaply and easily. While most alarms help operators detect and identify faults, some are unnecessary. A poor alarm system may cause alarm floods and nuisance alarms, which reduce the ability of operators to cope with plant abnormalities because critical alarms are buried in a lot of unnecessary alarms.

Since 1998, Idemitsu Kosan has actively worked on reducing the number of unnecessary alarms at its plants using the top-ten worst alarm method. This method collects data from event logs of alarms generated during operation and makes a list of frequently generated alarms. Then, the alarms are reviewed one by one, starting from the most frequently generated alarm, and the root causes that triggered them are identified. This method is effective for early stages of alarm reduction. In fact, average alarm frequency standards proposed by the EEMUA (2007) are achieved in some plants. However, when the proportion of the worst ten decreases, it becomes difficult to achieve effective improvements. For this reason, a novel alarm reduction method is required for further plant alarm reduction.

In this study, event correlation analysis was applied to alarm reduction for the Idemitsu Kosan ethylene plant in the Chiba complex. The event correlation analysis, originally proposed by Nishiguchi and Takai (2005), is a data-mining method that detects statistical similarities among alarms and operations. By grouping correlated

alarms and operations based on the degree of similarities, a strategy for reducing alarms can be found more easily than by analyzing individual alarms. This helps us to identify sequential alarms, repeated alarms caused by routine operations, and so on within a huge amount of alarm and operation data.

2. History of alarm reduction activities by Idemitsu Kosan

2.1. Ethylene plant in Chiba complex

Idemitsu Kosan started the ethylene plant at the Chiba complex in 1985. The ethylene plant is operated by two board operators using a DCS. The total numbers of tags in the DCS are 775 for process control and 2461 for process monitoring. The alarm data collection system was implemented at the ethylene plant in 2003. Figure 1 shows event generation frequencies of alarms and operations per board operator in the ethylene plant over ten minutes in 2003. In the ethylene plant, furnace decoking operations cause many repeated alarms.
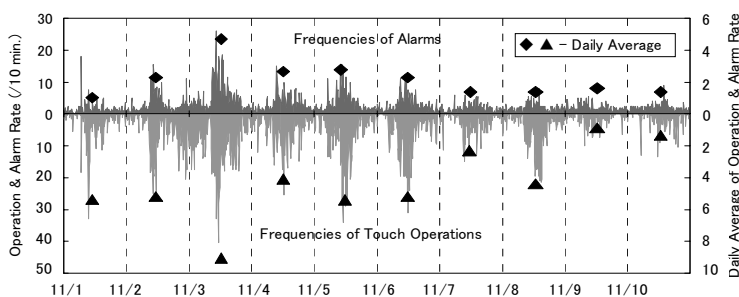
Fig. 1 Frequencies of alarms and operations in the ethylene plant

2.2. Alarm reduction by top-ten worst alarm method

Idemitsu Kosan applied the top-ten worst alarm method to the ethylene plant as a part of its total productive maintenance (TPM) activities in 2003. Figure 2 shows the ratio of the top ten alarms to all 16271 alarms generated in November of 2008 in the ethylene plant.
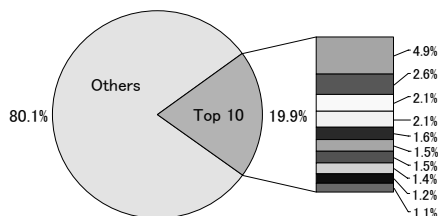
Fig. 2 Ratio of top ten alarms to all alarms generated in ethylene plant

We reviewed alarms one by one from the top of the list and reduced the number of alarms by changing alarm settings, retuning PID controller parameters, utilizing an alarm suppression system, and so on. Implementation of the programmable logic controller, where alarm settings are automatically changed according to plant states, drastically decreased a large number of alarms generated by unsteady state operations.

The top-ten worst alarm method was effective when the ratio of the top ten alarms was more than 20%. Figure 3 shows the transition of the frequency of alarm generation in the ethylene plant from 2003 to 2008. Though we continued the alarm reduction activities by the top-ten worst alarm method, the alarm generation frequency has not

been reduced since 2003. Because the ratio of each alarm in the top-ten worst alarm list became very small, it was difficult to achieve further effective improvement. Thus, a novel alarm reduction approach was desired to reduce the number of alarms.
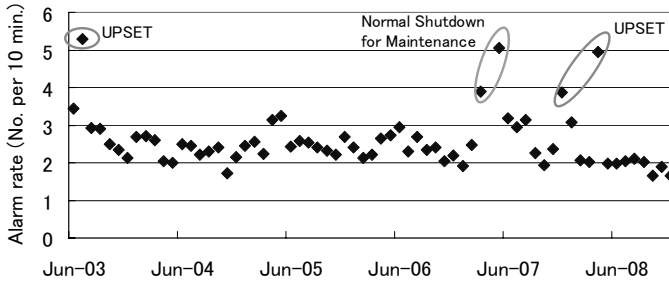
Fig.3 Average alarm rate in ethylene plant per board operator

3. Event correlation analysis

Event correlation analysis is a knowledge extraction (data mining) method intended event data composed of discrete events and the time they occur. This method uses event data from the plant to quantify the degree of similarity and time lag between two events by evaluating the cross correlation function. Then, the event tags are grouped using hierarchical clustering based on the degree of similarity.

3.1. Event log data

Plant data used in this method consists of occurrence time and tag name of an alarm or an operation, which we call an "event" hereafter. The plant data is converted into sequential event data $s_i(k)$. When event i occurs between $Ts + (k-1)\Delta t$ and $Ts + k\Delta t$, $s_i(k) = 1$. Here, Δt is the unit time, k denotes the discrete time, and Ts is the time of the first event. Figure 4 shows an example of the event data and the sequential event data $s_i(k)$.
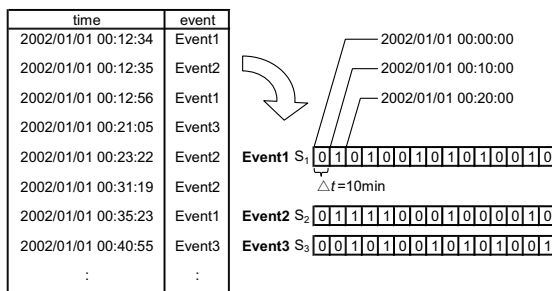
Fig. 4 Event log data

3.2. Cross correlation function

The cross correlation function between two sequential event data $s_1(k)$ and $s_2(k)$ for time lag m is defined by Eq. (1). Figure 5 shows the cross correlation function between two sequential event data. Here, K is the maximum time lag and T is the time period of the event data.

$$c_{12}(m) = \begin{cases} \sum_{n=1}^{T/\Delta t - m} s_1(n)s_2(n+m) & 0 \leq m \leq K \\ c_{21}(-m) & -K \leq m < 0 \end{cases} \tag{1}$$
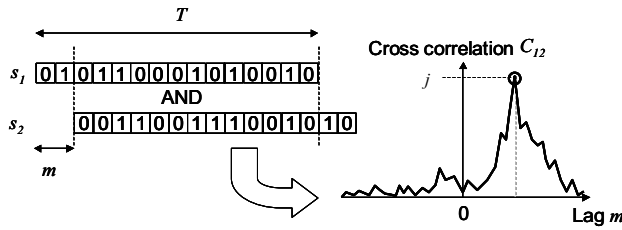
Fig. 5 Cross correlation function between two events

3.3. Definition of similarity

The probability of the occurrence of event i, which is denoted by p_i, is calculated by Eq. (2).

$$p_i \cong \frac{\sum_{k=1}^{T/\Delta t} s_i(k)}{T/\Delta t} \tag{2}$$

Assuming that any two events s_1 and s_2 are independent, the probability that two events occur simultaneously can be calculated by Eq. (3).

$$p_{12} = p_1 \times p_2 \tag{3}$$

Then, the probability distribution that two events occur simultaneously j times with time lag m can be expressed by the following binomial distribution. Here, j indicates the maximum value of $c_{12}(m)$.

$$P(c_{12}(m) = j) =_{T/\Delta t} C_j \, p_{12}^j (1 - p_{12})^{T/\Delta t - j} \tag{4}$$

If p_{12} is very small, Eq.(4) can be approximated by Poisson distribution.

$$P(c_{12}(m) = j) \cong \frac{\nu^j e^{-\nu}}{j!}, \qquad \nu = T/\Delta t \cdot p_{12} \tag{5}$$

The probability that two events occur simultaneously at least j times during time period T with time lag m is given by Eq. (6).

$$P(c_{12}(m) \geq j) \cong \sum_{i=j}^{T/\Delta t} \frac{\nu^i e^{-\nu}}{i!} \cong 1 - \sum_{i=0}^{j-1} \frac{\nu^i e^{-\nu}}{i!} \tag{6}$$

Therefore, the total probability that two independent events occur simultaneously more than j times with time lag m ($-K \leq m \leq K$) is given by Eq. (7).

$$P(c_{12}(m) \geq j \big| -K \leq m \leq K) \cong 1 - \left(\sum_{i=0}^{j-1} \frac{\nu^i e^{-\nu}}{i!} \right)^{2K+1} \tag{7}$$

Finally, the similarity S_{12} between two events s_1 and s_2 is calculated by Eq. (8).

$$S_{12} = 1 - P(c_{12}(m) \geq j \big| -K \leq m \leq K) \tag{8}$$

Larger similarity means stronger dependency or a closer relationship between the two events.

3.4. Grouping of events

After calculating similarities between all combinations of any two events in the plant log data, all events are classified into groups. By grouping events, it becomes possible to stratify and visualize the distance between events. In this study, the hierarchical clustering method was used.

3.5. Alarm reduction method

The following method is applicable to reduce unnecessary alarms and operations found by the event sequence analysis.
(1) Find a root cause of unnecessary alarms and operations based on the cause-effect analysis. Eliminating the root cause reduces unnecessary events.
(2) Find the first alarm in the sequential alarms, and retune settings of alarms following the first alarm. A state-based alarm is also effective.
(3) A programmable logic control is useful for reducing repeating alarms generated by routine operations.

4. Case study

To verify the effectiveness of the event correlation analysis method for reducing unnecessary alarms and operations, the analysis was applied to the Idemitsu Kosan ethylene plant in the Chiba complex. The plant log data gathered in one month included 1771 tags of alarms and operations. The total number of generated alarms and operations was 51640.

4.1. Results of grouping of alarms and operations

By applying the event correlation analysis to the plant log data, 1771 tags of alarms and operations were classified into 657 groups, where Δt, K, and the minimum threshold for identifying similarity between two events were set as 10 min., 30 min., and 0.95, respectively. Table 1 shows the result of the event correlation analysis. The total number of alarms and operations in the top 20 groups accounted for 45% of all events.

Table 1 Top 20 worst groups

No.	Group No.	Number of Events			Number of Tags		
		Total	Alarm	Operation	Total	Alarm	Operation
1	163	4198	0	4198	15	0	15
2	358	3570	3570	0	3	3	0
3	390	1314	0	1314	8	0	8
4	225	1223	83	1140	10	4	6
5	388	1158	47	1111	5	1	4
6	441	1072	141	931	13	7	6
7	179	1034	213	821	9	4	5
8	318	1003	0	1003	4	0	4
9	234	938	938	0	1	1	0
10	160	910	62	848	5	1	4
11	54	848	25	823	8	1	7
12	40	800	800	0	4	4	0
13	293	799	472	327	9	8	1
14	600	731	28	703	8	2	6
15	180	702	77	625	5	1	4
16	391	675	220	455	9	7	2
17	309	659	169	490	5	3	2
18	25	626	171	455	8	6	2
19	294	611	93	518	5	2	3
20	181	611	36	575	8	2	6

Figure 6 (a) shows 15 events and their relationships in the first group in Table 1. The total number of events in this group accounted for 8% of all generated events. According to an interview with the operators in charge, the events in this group occurred during furnace decoking. This means that 8% of the events could be reduced by improving the decoking operations. This indicates that the event correlation analysis is effective for reducing many alarms and operations.

Figure 6 (b) shows relationships among multiple alarms included in the seventeenth group in Table 1. Those alarms occurred in sequence. To reduce these sequential alarms, it may be effective to change their set points.

4.2. Identification of routine operations

When many operations are included in a group, these may be routine operations. Table 2 shows the points when alarms and operations occurred in the sixth group in Table 1. As shown in Table 2, alarms and operations in this group appeared synchronously. According to interviews with operators, these events occurred once or twice a week and were caused by routine operations. We automated these routine operations by a programmable logic controller. As a result, 70% of the alarms caused by these operations were eliminated.

(a) The top group (b) A group containing multiple alarms
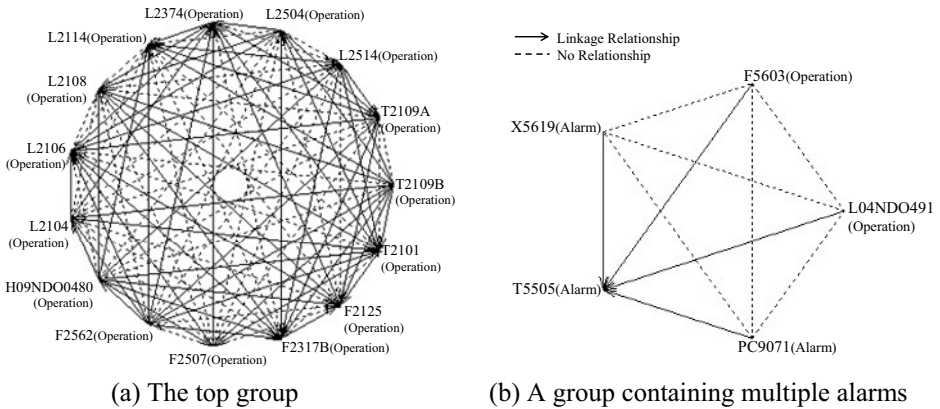
Fig. 6 Results of event correlation analysis

Table 2 Generation patterns of events in a group

5. Conclusion

Event correlation analysis was applied to the Idemitsu Kosan ethylene plant in Japan. Using this method, we effectively identified unnecessary alarms and operations within a lot of event data. This method is useful for reducing the number of alarms.

References

The Enginnering and Equipment Materials Users' Association (EEMUA), 2007, Alarm Systems A Guide to Design, Management and Procurement, EEMUA, London

J. Nishiguchi and H. Tsutsui, 2005, A New Approach to Process Alarm Reduction Using Statistical Point Processes, *SICE Annual conference 2005*, pp.443–448

T.Takai, 2007, Alarm Management in Chemical Plants, *Human Factors,* Vol.12, No.1, pp.10–23

Control Strategy with Distributed Action for Minimization of Transients in Distillation Column

Leandro O. Werle[a], Cintia Marangoni[a], Joel G. Teleken[a], Claudia Sayer[a], Ricardo F. Machado[a*]

[a]*Santa Catarina Federal University, Chem. Eng. Dep. University Campus, Mail Box: 476, Florianópolis-SC, Brazil, Zip Code: 88010-970 *Email: machado@enq.ufsc.br*

Abstract

Distillation columns require well adjusted control systems to quickly reject disturbances. To assist this purpose, a new distributed control strategy using electrical resistances on intermediate plates of a distillation column was proposed. Usually the control of distillation columns is accomplished centralized at the bottom and at the top. In the present work the effect of the distributed control is evaluated, through the addition of thermal energy using an electric resistance located at an intermediate plate, when a disturbance is applied in the composition of the feeding stream. In order to validate the proposed strategy, a pilot distillation unit composed of 13 trays was used, processing an ethanol-water mixture with a feed flow rate of 300 L/h. The objective of this paper is a reduction in the operation transients. In addition, a thermal balance was used to verify if distribution of thermal energy occurs between the bottom of column and the plate where the distributed control is applied. The results showed that the introduction of the distributed heating along the column allows faster dynamics, representing a valid option for the reduction of transients.

Keywords: distillation column, distributed action, minimization of transient.

1. Introduction

The distillation process is widely used in industries, especially in petroleum refineries. The dynamic behavior of the process is intrinsically non linear and variables are highly coupled, being an essentially multivariable problem and with operation restrictions (Miladi and Matjaba, 2004).

Distillation columns require well adjusted control systems to quickly reject disturbances. However, it is known that well designed and adjusted control systems are not enough to eliminate the operation transients. The transient formation in a distillation column occurs when the process is disturbed or when an external factor induces the modification of the operation point of the unit. In the first case there are factors such as variable coupling, nonlinearities, time-delay, high time constants and process constrains. Whereas the second case involves aspects like the mixture to be distillated, feed composition changes and operation transitions due to market alterations.

Procedures of campaign changes are problematic in distillations, requiring to drive the system to a new operation regime as fast as possible to minimize the formation of out of specification products. Although several researches involve quite complex control approaches, only few works approach the matter of transient minimization. Usually, the distillation column control is carried out in a centralized way in the variables on the bottom, top and lateral outlets, generating a high transition time when the process is

disturbed. In previous works (Marangoni et al., 2005) a distributed control strategy was proposed, by using electrical resistances as an energy source on the trays.

It is known that in the operation of distillation columns the largest cost is in the amount of applied energy. In this study we chose to work with electrical resistance and a plate heat exchanger (reboiler). It should be emphasized that, instead of conducting the heat through the electrical resistance, it could do so through the use of heat exchangers or condensers, to perform the heat exchange. According to Agrawal and Herron (1998) the appropriate use of reboiler and intermediate heat exchangers can promote significant reductions in the cost of operating of a system of distillation. The authors perform a detailed analysis of some configurations using such equipment.

In addition, several studies involve heating at intermediate trays from the point of view of diabatic distillation (Rivero, 2002), aiming energetic saving. Some columns in petroleum refineries use circulating refluxes in order to promote a better separation. Nevertheless, in both of these applications intermediate heating is not used as part of the control loop. This paper analyzes the behavior of a distillation column when a disturbance is applied in the composition of the feed stream using two different approaches: conventional and distributed control strategy. A thermal balance was used to verify if in the second approach the thermal energy added to the column is distributed between the bottom of column and the plate where the new control strategy is applied.

2. Methodology

The experiments were carried out in a pilot distillation unit, composed of 13 trays, instrumented with a fieldbus digital communication protocol, processing an ethanol-water mixture with a feed flow rate of 300 L/h.

2.1. Experimental Unit

The pilot unit operates continuously, with the feed stream inserted in the 4th tray, with a total height of 2.70m, built in modules with 0.15m height and 0.20m diameter.

2.2. Control System

The control configuration of the distillation column was formulated based on Nooraii et al (1999) research. The following control loops were defined: (1) bottom level control through the bottom product flow rate adjustment; (2) reflux accumulator level control by manipulating the top product flow rate; (3) feed flow rate control as a function of the adjustment of the same stream flow rate; (4) feed temperature control through the fluid flow rate adjustment in the heat exchanger of this stage; (5) last tray (distillate) temperature control by means of the manipulation of the reflux flow rate; (6) reboiler temperature control through the vapor flow rate in the heat exchanger of this stage and (7) temperature control of pre-defined stages of the column through the adjustment of the dissipated power in the tray electrical resistance.

The first, second and third loops represent the column mass balance (inventory) control. The fifth and sixth loops comprise the quality control – in this case represented by the temperature. When these two loops are used combined we called it conventional control. When these two loops are combined with the seventh loop described above, we consider it as the distributed strategy.

PID controllers were adjusted for the temperatures of the bottom and of the last tray, characterizing the conventional control. The controllers used for the bottom level and the feeding flow rate consist of a cascade system composed by a feedback PI and another anticipative (feedforward). To adjust the controllers the classical methods of Cohen-Coon, ITAE and Zieger-Nichols (Seborg et al., 1989) were applied. The distributed control was carried out by using an intermediate tray, tray 2.

2.3. Experimental Procedure

For this study disturbances were introduced in the feed composition. The performance of the distributed approach (control on the bottom, top and tray 2) was compared with the conventional configuration. Experiments were conducted with initial volumetric compositions of ethylic alcohol in the feed of 10% (v/v), the feed temperature was controlled at 92ºC, the feed flow rate was 300 L/h and the pressures on the bottom and top varied around 1.25 and 0.25 bar, respectively. Composition measurements were carried out during the experiments. The distributed heating was carried out by means of electrical resistances with maximal power of 3.5 kW each. For the identification of the most sensitive stage for the application of the distributed control sensibility analysis methods were used: successive plates, sensitivity symmetry and maximum sensitivity, as detailed in Marangoni (2005). The sensibility analysis was accomplished with base on the studies of Luyben (2006). The results showed that in the stripping section plates 2 and 3 were the most sensitive plates, and in the rectifying section trays was 5 and 7. Plates 3 and 5 are adjacent to the feeding plate, receiving disturbances directly in their flow rate, for this reason these plates were not selected for the distributed heating. Between the remaining "most sensitive" plates (2 and 7), plate 2 was chosen for application of the distributed control, since the stripping section is the region that demands more energy. This decision is also based on the operation mode of diabatic distillation columns in which the rectifying section receives heat, and the rectification section releases heat through heat exchangers. The choice of using a single plate for the distributed heating also aimed to avoid more than one simultaneous effects.

2.4. Energy Balances

An energy balance, shown in Equation 1, was used to calculate the heat inserted by the reboiler. Assuming that the losses to the environment are negligible, on the steady state:

$$Q_{reb} = (Q_c + Q_d + Q_b) - (Q_{res} + Q_f) \tag{1}$$

Where: Q_{reb} is the heat inserted into the bottom by the reboiler, Q_c is the heat withdrawn by the condenser, Q_d and Q_b are, respectively, the heats that come out of the column by the destilate and bottom product streams, Q_{res} is the heat added by the electrical resistance and Q_f is the heat of the feed stream, all in [kW].

3. Results

The disturbance applied in the feed composition was a positive step, the volumetric fraction of ethanol varied from 10% to 20% (v/v), with the addition of ethanol to the feed reservoir. Due to the largest volume of the reservoir the contributions of the streams composition of the top and bottom that return to the feed reservoir are small (negligible) in the composition of the feeding. It is important to mention that in this paper the reference values for the control loops were maintained after the introduction of the disturbance.

Figure 1a shows the effect of the disturbance on the control loop of the temperature of the reboiler. It can be observed that after the control system assimilates the disturbance, a temperature decrease occurs in both configurations, however the distributed control resulted in a lower decrease (0.5ºC), when compared to the conventional one (1.2ºC), this is explained by the contribution of the resistance to the formation the vapor phase.

In the same way as in the stripping section, the rectifying section is also affected by the disturbance of the feed composition. This effect can be observed in Figure 1b, the increase of the top temperature, observed for both configurations, is due to the thermal load added in excess by the reboiler in order to keep the reference values, since the operation points were not adjusted after the disturbance. The disturbance affected the

liquid-vapor equilibrium, leading the vapor to drag great amounts of water to the top of the column. This resulted in temperature increases for both approaches, due to the saturation of the reflux valve in both configurations (Figure 1b).
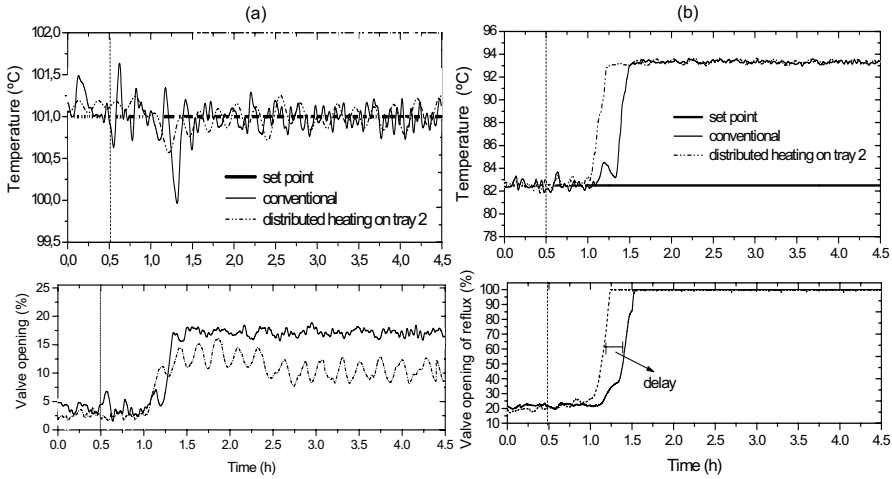
Figure 1 - Response of the temperature control loops: a) reboiler and b) last stage.

In the results shown in Figure 2, temperature of tray 2, it might be observed that 30 minutes after the application of the disturbance the temperature of tray 2 starts to decrease for the conventional configuration, due to the increase of the amount of the more volatile component.

In the distributed configuration, on the other hand, this decrease was not observed due to the instantaneous energy supply by the electrical resistance. The energy provided by resistance R2 to the system, increasing the vapor phase, may have favored the transport of the vapors rich in water to the top of the column, explaining the premature saturation of the reflux valve of the temperature control loop of the last stage (Figure 1b) and the dynamic differences between both configurations.
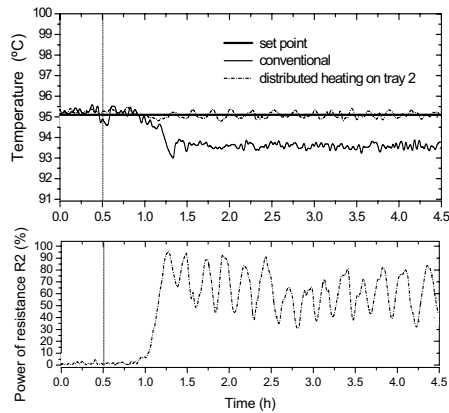
Figure 2 - Effect of the disturbance in the temperature control of tray 2.

Comparing these results with those of Marangoni (2005) a similar behavior might be observed for these three control loops analyzed when a disturbance was introduced in

the composition of the feed stream. Finally, the analysis of the behavior of the manipulated and controlled variables in Figures 1 and 2, shows that the introduction of a heating point distributed along the column improves considerably the performance of control loops.

The disturbance applied volumetric fractions of ethanol in feed stream, affects the fraction of this component in the stripping section before the rectifying section. After the alteration of the composition at the bottom the change is propagated through the vapor phase to the upper trays, establishing a new composition at the top, which in turn affects the composition at the bottom through the liquid that flows down the column. The equilibrium between the phases is reached after a few hours.

Figure 3 shows the derivative of the volumetric fraction of ethanol in the top. These results show that the distributed control is faster (around 1 hour) than the conventional control, mainly due to its different dynamics.

The high temperature reached at the top of the column led to very high reflux flow rates (Figure 1b), which in turn ceased the production of distillate. Therefore, the new steady state reached after the disturbance cannot be considered appropriate for the operation since it does not offer conditions for an adequate separation of the components. Showing, this way, that this kind of disturbance needs a new adjustment of the process operation point. According to Hurowitz et al. (2003), disturbances in the feed composition consist in the main challenge for distillation column dynamics.
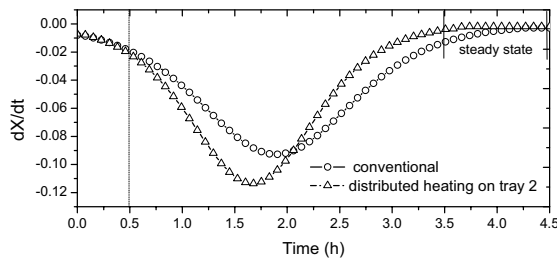
Figure 3 - Derivative of the volumetric fraction of ethanol in the top product.

The effect of this disturbance, increase of the fraction of the more volatile component in the feed stream, which leads to an increase of the fraction of this component in the bottom section, can be compared to the effect of a positive disturbance on the thermal load of the reboiler, leading to an increase of the vapor flux throughout the whole column, vaporization of the less volatile components and raise of the their fractions in the top, coupled with a temperature increase throughout the whole column.

The thermal energies of the streams, calculated by the global energy balance (Equation 1) are shown in Table 1. Comparing the absolute values of the added energies it might be observed that in the distributed control the energy added in the bottom in the conventional control is distributed between the bottom and the resistance of tray 2.

In the Figure 1a it might be observed that after the disturbance the bottom valve opening operated with smaller values in the distributed control compared to the conventional one, compensating this way, the energy added in plate 2. In the conventional approach, after the disturbance, the overall heat supply of reboiler increases 45.41 kW, while in the distributed approach the increase was only 43.34 kW, and 2.28 kW were added by the electrical resistances, showing this way the distribution of overall heat supply. This distribution of heat was also observed by Werle et al. (2007) when used the distributed control, applying a different type of disturbance (in the feeding flow rate).

Table 1 - Thermal energies (kW) calculated at steady state.

Conventional	Q_{reb}	Q_f	Q_{res}	Q_c	Q_d	Q_b
Before	9.80	21.71	----	9.75	0.34	21.41
After	55.21	20.68	----	55.17	1.11	19.61
Distributed healting						
Before	9.70	21.68	0.05	9.61	0.40	21.42
After	53.04	20.68	2.28	55.49	1.15	19.54

The great increase of the thermal energy of the reboiler, after the disturbance for both studied control configurations is a result of the increase of the ethanol concentration, keeping the feed stream temperature constant.

4. Conclusion

The distributed control approach led to a reduction of approximately 1 hour of the operation transient when compared to the conventional process. In addition, the distributed control also resulted in a less oscillatory behavior in most of the control loops. Both these advantages observed when the distributed control was used were obtained with nearly the same overall heat supply of the conventional approach, distributed along the column. Thus, the introduction of the distributed heating along the column has shown itself as a valid option for the transient reduction.

5. Acknowledgments

References

A. Noorai, J. Romagnoli, J. Figueroa, 1999, Process, identification, uncertainty characterisation and robustness analysis of a pilot scale distillation column. Journal of Process Control, v. 9, n 3, p. 247-264.

C. Marangoni, 2005, Implementação de uma Estratégia de Controle com Ação Distribuída em uma Coluna de Destilação. Tesis. Programa de Pós-Graduação do Dep. de Eng. Química e Alimentos, Univ. Federal de Santa Catarina, Florianópolis.

D. E. Seborg, T. F. Edgar, D. A. Mellicamp, 1989, Process Dynamics and Control, John Wiley & Sons, Singapore.

L. O. Werle, C. Marangoni, F. R. Steinmacher, R. C. Soares, P. H. H. Araújo, R. A. F Machado, C. Sayer, A. Bolzan, 2007, Uso de aquecimento distribuído em uma coluna de destilação: efeito da perturbação na vazão de alimentação. In: 4° Congresso Brasileiro de P&D em Petróleo e Gás - PDPetro.

M. M. Miladi, J. Mutjaba, 2004, Optimisation of design and operation policies of binary batch distillation with fixed product demand, Comp.&Chem. Eng. v.15, p.2377-2390

R. Rivero, 2002, Application of the exergy concept in the petroleum refining and petrochemical industry. Energy Conversion and Management, v. 43, p. 1199-1220.

R. Agraval, D.M. Herron, 1998, Intermediate reboiler and condenser arrangement for binary distillation column. AICHE Journal, v. 44, p. 1316-1324.

S. Hurowitz, J. Anderson, M. Duvall, J.B. Riggs, 2003, Distillation Control Configuration Selection, Journal of Process, v. 13, p. 357-365.

W. L. Luyben, 2006, Evaluation of criteria for selecting temperature control trays in distillation columns. J. Proc. Cont. v.16, p.115-134.

1533

Experimental Startup of a Distillation Column Using New Proposal of Distributed Heating for Reducing Transients

Leandro O. Werle[a], Cintia Marangoni[a], Joel G. Teleken[a], Claudia Sayer[a], Ricardo F. Machado[a*]

[a]*Santa Catarina Federal University, Chemical Eng. Dep. University Campus, Mail Box: 476, Florianópolis-SC, Brazil, Zip Code: 88010-970 *Email: machado@enq.ufsc.br*

Abstract

This paper proposes the use of distributed heating along the trays of a distillation column, which is important during transient periods, particularly during startup. The experiments were carried out in a pilot distillation column with 13 sieve trays processing a binary mixture composed of water and ethanol. It was verified that the use of a heating system that is able to act in a distributed way allows a reduction in the time required for the startup procedure. This process was chosen due to its characteristics of inertia, non-linear behavior, and long transient periods, which make the objective an even greater challenge since this process is hard to control with traditional methods. Most of the previously published studies on startup involve simulations, whereas the main contribution of this paper is that is provides experimental results.

Keywords: distillation column, startup, distributed heating.

1. Introduction

Distillation columns have been extensively studied and this type of research could be considered as complete (Eden et al., 2000). However, the control of a distillation column can be split into two different problems: steady state and dynamic control. Dynamic control is necessary when the column is operating out of its normal operation conditions, which generates strong nonlinearities in the process. Typical situations are the startup and shutdown operations, in which the plant is far from under its normal production conditions, and therefore these may display very disparate behavior.

The startup of a continuous distillation column presents a series of operational problems, as described by Kister (1979), which until now have been difficult to minimize. Most of the proposals address controlling this process phase. Fabro (2005) proposed the use of intelligent control techniques such as a neural networks, fuzzy systems and genetic algorithms for distillation column startups, allowing the desired steady state to be reached as quickly as possible. Along the same research lines a new on-line control algorithm has been proposed by Barolo et al. (1993), based on Generic Model Control, for improving the automatic startup of a binary distillation column. Scenna and Benz (2003) described multiple steady states for different initial conditions as starting points. As can be observed, several methods and configurations for distillation column startup have been proposed in the literature by different authors. In this study we try to make use of a new control strategy initially proposed by Marangoni (2005), which uses a distributed corrective action based on heating the distillation column trays, and evaluate this new configuration during the startup process. The startup of a distillation column can be considered in three distinct steps: the heating of

all stages, the introducing of a reflux stream in manual mode, and the changing of the control loops to automatic mode. The main goal of this study is to analyze these three steps of the column startup, which need special care and require enormous amounts of energy.

2. Methodology

The experiments were carried out in a pilot distillation unit, composed of 13 trays, instrumented with a digital fieldbus communication protocol, processing an ethanol-water mixture (sub-cooled liquid) with a feed flow rate of 300 L/h.

2.1. Experimental Unit

The pilot distillation unit operates continuously, with the feed stream inserted in the 4th tray, with a total height of 2.70 m, built in modules with 0.15 m height and 0.20 m diameter. A schematic diagram of the unit is shown in Figure 1.
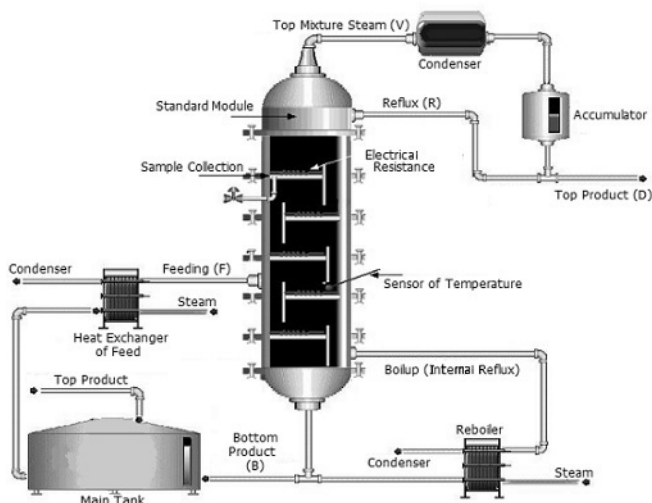
Figure 1 - General scheme of the equipment and pilot distillation unit.

The main feature of this distillation unit, which makes it different from the others, is the use of heating points supplied by electrical resistances distributed along the column, which decentralize the heat supply from the bottom.

2.2. Experimental Procedure

In the conventional approach, the column is operated with heat supply only at the reboiler, with the manual mode of the vapor valve opening fixed at 8% (20kW), as represented in Figure 3b. When the proposed configuration is applied, the heat is supplied at the bottom by the reboiler (with the same previous overall heat supply) and at an intermediate point, by way of the electrical resistance on the 2nd tray, started with 50% of its capacity (1.75kW) The procedure adopted for the column startup in the conventional operation is based on a study by Steinmacher et al. (2004). In this study the startup of the distillation column was considered to have three distinct steps: the heating of all stages, the introducing of the reflux stream in manual mode, and the changing of the control loops to automatic mode. In the first step, at the beginning of the procedure, volumetric compositions of ethanol in the feed and accumulator tanks

(partially filled with mixtures in the desired compositions) were 15 and 80% respectively. During the startup, the feed stream is continuously introduced and controlled at 300 L.h^{-1}. To do this experimentally the process is closed, that is, the bottom and top product streams form the feed stream in the main tank (Figure 1). Thus, the startup is initiated with the withdrawal of bottom product (bottom level control). The liquid mixture from the feed descends throughout the column until it reaches the bottom, where the heat exchanger heats and vaporizes this stream. The vapor mixture ascends through the column, heating it tray-by-tray until it reaches the condenser. Here step 1 is finalized. The condensed mixture is received by the accumulator, which is already filled with the material at the desired composition (80% v/v ethanol) for the reflux stream. The reflux is total until a stable situation is achieved (temperature deviation is null over time) which characterizes the end of step 2. At this moment, the top product withdrawal is started and the accumulator level bottom and top temperature control loops are changed to automatic mode, starting step 3.

The steady state is determined when the accumulator temperature and top composition do not vary over time. These variables were used as evaluation parameters. Volumetric composition measurements corroborate this analysis. The composition was measured with a densimeter for alcohol.

The startup procedure with the distributed heating approach was almost the same. The only difference was the use of an electrical resistance with its total power (3.5 kW) at one intermediate plate, enabled from the beginning of the experiment.

For the identification of the most sensitive stage for the application of the distributed control sensitivity analysis methods were used: successive plates, sensitivity symmetry and maximum sensitivity, as detailed in Marangoni (2005). The results showed that in the stripping section plates 2 and 3 were the most sensitive plates. In this study, plate 2 was chosen for application of the distributed control to avoid more than one simultaneous effect. Rectifying trays was not used because at the beginning of startup only the stripping section has enough liquid content in the trays to allow heating through electrical resistance.

3. Results

The temperature profiles of a stripping tray (tray 3) and a rectifying tray (tray 13) are represented in Figure 2. In these experiments a difference was noted in the tray behavior along the sections of the column during the heating. The *stripping* trays show a linear heating region, and which is not observed in the *rectifying* trays. Tray 3 was chosen to illustrate the stripping section as it is the tray which best represents the heat addition effect. This stage is successive to tray 4 which receives the feed stream, as also indicated by the sensitivity analysis.

The results shown in Figure 2 indicate that before the beginning of the reflux, the temperature reached at the top of the column is slightly higher for the distributed configuration than for the conventional one. However, the temperatures after the introduction of reflux were almost the same in both situations. This confirms that the steady state and the reflux flow rate (30L/h) were the same in both cases.

After the linear region (linear phase), shown in Figure 2, an interval can be observed. This is the time required to heat the tray contents. In this step, the heating is slow until the bubble point temperature of the mixture is reached, and the tray temperature increases rapidly. When using the distributed strategy, the time saving is the same for both the third and thirteenth trays (0.32h), when compared with the conventional strategy.
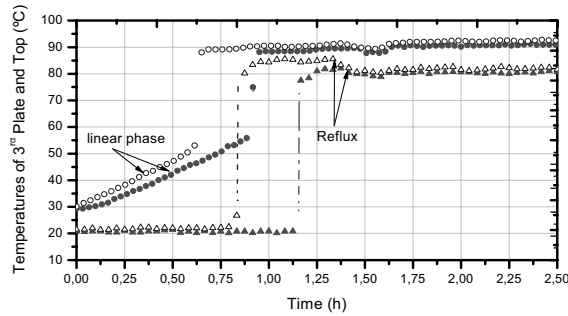
Figure 2 - Temperature profiles of the third tray: conventional (●) and distributed heating applied to tray 2 (○); and of the thirteenth tray, conventional (▲) and distributed heating applied to tray 2 (△), during the process startup.

An analysis of the angular coefficients of the linear regression in the feed and tray 3 temperatures for conventional and distributed modes was carried out. The results are summarized in Table 1. In all of the curves the multiple regression coefficient obtained was higher than 0.997. It is observed that, for both temperatures, angular coefficients are higher when the distributed strategy is used. These results, along with those shown in Figure 2, indicate that the distributed heating resulted in a better performance, a faster action and a reduction in the time required to reach the desired temperature at the top the column.

Table 1 - Coefficients of the linear phase of the heating at the temperatures of the feed stream and of tray 3.

Model	Linear Regressiom	
	Conventional	Distributed Heating
Feed temperature	Y=34.090.X + 31.235	Y=40.517.X + 31.527
Tray 3	Y=31.532.X + 29.761	Y=35.155.X + 30.153

Figure 3 shows the profile of the bottom temperature before and after the control loop is changed to automatic mode with the two studied configurations. The distributed heating again showed a faster action when compared with the conventional process. Using the intermediate heating, the desired temperature value (98°C) was reached 0.75h faster than using the conventional configuration. This is of particular interest as it shows that the action of the resistance increased the heating rate of the bottom temperature.

In order to better evaluate the effect of the distributed heating, the evolution of the derivatives of the bottom temperatures were compared for both experiments. These derivatives were obtained from sigmoidal functions, adjusted to describe the bottom temperature in each of the cases, as shown in Figure 3c. It was verified that the proposed distributed configuration allowed the steady state to be reached faster. A period of 1.4h was required for the bottom temperature to become stable with the conventional configuration and only 0.8h for the distributed configuration. This represents only 57% of the time required to reach the desired bottom temperature.

In addition, in Figure 3b it can be observed that after changing the control loop to automatic mode the bottom valve operated with smaller openings, thus compensating the energy added to plate 2.
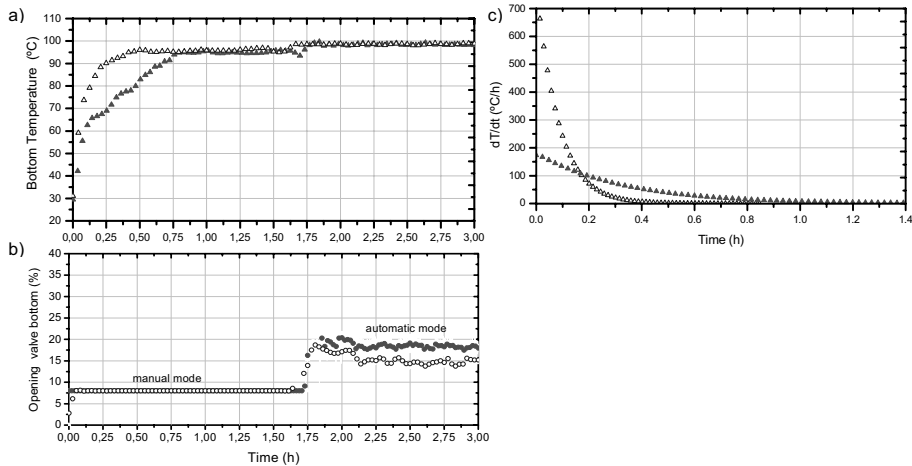
Figure 3 – a) Column bottom temperature, conventional (▲) and distributed heating applied to tray 2 (△), b) manipulated variable conventional (●) and distributed heating applied to tray 2 (○) and c) derivatives of the bottom temperatures as a function of time.

The reduction in the time required to reach the steady state when the distributed heating is applied is confirmed by the measurements of the volumetric ethanol fractions in the bottom column stream and their derivatives (dX/dt) to characterize the steady state, as shown in Figures 4a and 4b, respectively.
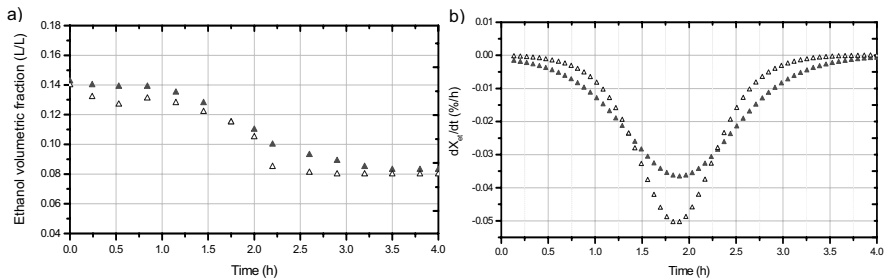
Figure 4 - a) Volumetric ethanol fractions in the bottom stream of the column, b) Derivatives of the volumetric ethanol fraction as a function of time, conventional (▲) and with distributed heating applied to tray 2 (△).

It can be observed that there is a change in the bottom ethanol fraction which starts at 0.14% and ends at 0.08%. This is due to the introduction of reflux. Through this observation it is possible to analyze the period out of steady state - as illustrated by the derivative of this variable in relation to time. Is important to note that reflux is introduced in the 13th stage – more distant from the bottom. This means that the interaction between the bottom and top is responsible for the longer time required to reach steady state (or to maintain the column out this condition). As illustrated by other results, Figure 4b shows that the distributed configuration allows that the transition period is shorter than the conventional one (around 0.75h).

In Figure 5 the results obtained in terms of the time required for each column startup step are summarized. The effect of the distributed heating in relation to the conventional heating can be clearly observed, particularly during steps 1 and 3.
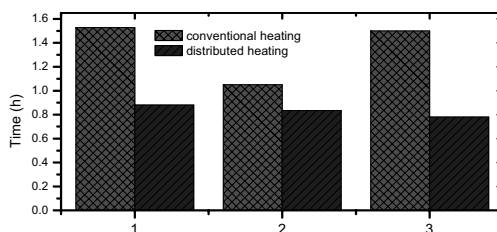
Figure 5 - Startup steps: 1 Heating all stages, 2 Manual Reflux, and 3 Reflux + Bottom product in the automatic mode.

In all three stages the reduction in the time, due to the effect of the addition of the intermediate heating point, was significant when compared to the conventional procedure, being in the order of 43%, 21% and 50%, respectively. Thus, a reduction of approximately 40% in the total startup time was achieved in the distillation column operating with heat distribution, which suggests considerable gains in the process.

4. Conclusions

Comparison between the two operation approaches, showed that the temperatures of the trays of the stripping section increase at a higher rate when the distributed heating with a resistance on plate 2 is used, reaching a permanent regime in a much shorter time than in the conventional process. Thus, the introduction of distributed heating along the column is demonstrated to be a valid option for reducing the startup time. In this study, a reduction of approximately 40% in the startup time was observed in relation to the conventional process, allowing faster dynamics and lower operation costs, with almost the same amount of energy as the conventional configuration. The reduction in the time required for startup, resulting in a considerable decrease in the production of out-of-specification products, compensates for the slight increase in the energy supply.

5. Acknowledgments

References

C. Marangoni, 2005, Implementação de uma Estratégia de Controle com Ação Distribuída em uma Coluna de Destilação. Tesis. Programa de Pós-Graduação do Dep. de Eng. Química e Alimentos, Univ. Federal de Santa Catarina, Florianópolis.

F. R. Steinmacher, C. Marangoni, R.F.A Machado, P. H. Araújo, 2004, Startup evaluation and definition of operational conditions of a distillation column. Annals of the Brazilian Congress of Petroleum and Gas.

H. Z. Kister, 1979, When tower startup has problems - Hydrocarbon Processing. Chem. Eng. Science, v. 2, p. 89–94.

J. A. Fabro, L.V.R. Arruda, F. Neves Jr., 2005, Startup of a distillation column using intelligent control techniques, Comp. & Chem. Eng., v.30, p.309–320.

M. Eden, A. Koggersbøl, S. B. Jørgensen, 2000, Dynamics and control during startup of heat integrated distillation column, Comp. & Chem. Eng., v. 24, p. 1091–1097.

M. Barolo, G. B. Guarise, S. Rienzi, A. Trotta, 1993, On-line startup of a distillation column using generic model control. Comp. & Chem. Eng., v. 17, p. 349-354.

N. J. Scenna, S. J. Benz, 2003, Start-up operation of reactive columns with mulitple steady states: ethylene glycol case, Ind. Eng. and Chem. Research, v.42, p.873-882.

Failure Diagnostics Using Data Mining Tools

Carlos A. Vaz Jr.,[a] Ofélia de Q. F. Araújo,[a] José Luiz de Medeiros[a]

[a]*Departament of Chemical Engineering, Escola de Química da Universidade Federal do Rio de Janeiro. Sala E-201, Bloco E, Centro de Tecnologia, Ilha do Fundão, CEP: 21949-900 Rio de Janeiro, Brazil*

Abstract

The increase in the use of electronic instrumentation on the chemical industry produces a large operational database. Several methodologies have been applied to the task of review and extract relevant information from these data. A software based on principal component analysis (PCA), hierarchical classifiers and prototypes method was developed. This software detects operational failures using time series generated by the process. The process is an industrial pipeline network for naphtha transport. Two different failures were analyzed: sensors failures and leaks. The software presents a satisfactory failure detection performance. After the leak was detected, the software must proceed to locate and quantify it. The leak parameters – its diameter and location – were estimated correctly for severe failures. The success of detection proceeds is strongly influenced by the severity of the problem. This successful rate reaches more than 95% for severe failures.

Keywords: pipeline, leak detection, sensor failure detection, data mining

1. Introduction

The transport of liquids by pipelines is applied to the displacement of several products, such as water, oil, gas and fuel derivatives. Although highly reliable, pipelines may present severe accidents in some situations, like fatigue and collapse of the tube material. Pipes transporting toxic and/or flammable products are inherently associated with a high potential risk factor (Silva et al., 1996).

Pipelines used for a long time with both high pressure and flow rate may suffer material fatigue and collapse. This situation produces leaks of material and environmental damage, loss of inventory, fires, explosions and the possibility of loss of human lives. Preventing leaks is very important to the safe operation of a pipeline and so are procedures to minimize the extent of its consequences in case of a leak occur. Then, leaks become a constant concern of the chemical industry, regulatory agencies, and society.

In addition to leaks, another kind of pipeline's disruption is the failures in sensors. The development of micro-electronics and its industrial applications create low expensive equipments. Thus, the number of sensors in use in many different industry areas has grown strongly, increasing the possibility of failure in any of these electronic components (Singhal, A., Seborg, D., 2002).

Identifying sensors failures is important for the correct operation of complex pipelines networks. Many methods for detection of anomalies are based on comparison of real time process data and predictions generated by a mathematical model. The fail is detected if significant differences between the process values and model predictions are observed. The models applied to produce process predictions can be

"phenomenological" or "black box". "Black box" models do not require physical description of the process. Examples of "black-box" models are "neural networks", "AR" (autoregressive models) and "ARMA" (autoregressive moving average) (Vandaele, W., 1983).

In this article, tools based on principal component analysis (PCA), hierarchical classification and prototypes were applied to detect, localize, quantify and identify failures.

2. Methodology

This item describes the mathematical methods used in this work. The methods presented were applied to allow detect the presence of sensors failures or leaks. The methodologies also identify the type, origin and severity of the anomaly.

2.1. Hierarchical Classification

Hierarchical classifiers try to separate or classify observations contained in a wide range of data. Groups are formatted to promote the separation. The members of each group show high similarity between them. Similar observations are placed in the same group, and "non-similar" objects are placed in different groups (Hart P., Duda, R. O., 2000). Hierarchical classifier was applied to identify observations "with" and "without" the presence of failure. The similarity between two observations can be understood as inversely proportional to spatial distance between these points. Or, the spatial distance between two points are proportional to the "dissimilarity" between the observations.

In a two dimensional space (R^2), the distance between point "i" and point "j" in the plane is traditionally defined using Euclidean metric. This metric can be extended to a generic spatial dimension "n" (R^n) using Equation 1.

$$d_{ij} = \left[\sum_{k=1}^{n} \left(x_{i,k} - x_{jk} \right)^2 \right]^{1/2} \quad (1)$$

The Euclidean distance is the most applied distance metric, but it is not the only definition for "distance" between two points. The "City Block" metric (Equation 2) and "Minkowski" (Equation 3), described by Duda and Hart (2000) can be applied too. In some situations, City Block and Minkowski show greater effectiveness in the characterization of similarity between two observations. The City Block and Euclidean are special cases of Minkowski equation, using "r" equal to 1 and 2 in Equation 3.

$$d_{ij} = \sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right| \quad (2)$$

$$d_{ij} = \left(\sum_{k=1}^{n} \left| x_{ik} - x_{jk} \right|^r \right)^{1/r} \quad (3)$$

Once defined the metric classification, the classification procedure is simple. A matrix of distances between all the data observations is created. Then, the two points with smaller distance between them are selected. These two observations are merged, creating a new point. A new matrix is prepared, and the procedure is repeated. The procedure continues until all data is allocated in two clusters. The observations on the same group must present high similarity between them, and strong dissimilarity when compared to data of another cluster. Finally, the identification of the two groups

("failure" and "without failure") is promoted classifying an observation previously known. Besides allowing the detection of failures, the hierarchical classifier also quantifies the severity of the fault.

2.2. Method of Prototypes

One approach to identification and location of the faults was based on method of prototypes. To implement this methodology, representative cases of major anomalies are selected, and these observations are called "prototypes". The prototype must be appropriated to identify similar situations.

After an operational anomaly has been detected, the experimental data of this failure is then compared with the prototypes, looking for an elevated similarity. A PCA model is adjusted for each prototype and for the experimental data. The models are compared using a similarity methodology called "Distance Similarity Factor" (Sdist). The prototype identification allows estimate the "type" (leak or sensors failures) and the "location" of the fault. The PCA and Sdist methodologies are described below.

2.3. PCA Methodologies

The principal component analysis (PCA) methodology can be described as the calculation of eigenvectors and eigenvalues of a covariance matrix obtained from the normalized data. The eigenvectors, especially those associated with larger eigenvalues; provide important information about the pattern of data distribution (Hart P., Duda, R. O., 2000).

The eigenvalues define the amount of information described in each eigenvectors. This article used eigenvectors enough to describe 99% of the pattern of data distribution. The eigenvectors are called "components", while the eigenvectors associated with the high eigenvalues are called "principal components".

The matrix of components (MC) is composed by the principal components, and the PCA model is represent by matrix "C" (equation 4).

$$C = MC * MCT \qquad (4)$$

The matrix "C" is multiplied by experimental data ($\hat{x}\exp_k$) resulting in the modeled data ($\hat{x}\mod_k$) (equation 5).

$$\hat{x}\mod_k = C \; \hat{x}\exp_k^T \qquad (5)$$

2.4. Distance Similarity Factor (Sdist)

The Sdist factor is the probability that the center of the data set A ($\overline{x}_A$) is at distance ϕ (or less) from the center of the data B (Singhal, A., Seborg, D., 2002). The Sdist is calculated using equation 6, where ϕ is the Mahalanobis distance.

$$Sdist = 2\left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\phi} e^{-\frac{z^2}{2}} dz \right] \quad , \qquad \phi = \sqrt{(\overline{x}_A - \overline{x}_B)\sum_s^{*-1}(\overline{x}_A - \overline{x}_B)^T} \qquad (6)$$

2.5. Pipeline Network

To test the methodologies proposed here, a pipeline network simulator developed by Vaz Junior (2006) was employed. The computer simulation allows the realization of a large number of experiments and tests involving the pipeline, with different conditions of sensor faults and leaks.

The pipeline network employed here transport naphtha over a length of 30 km connecting the vertex "1" (origin) to the vertex "8" (destination). A network diagram is shown in figure 1.
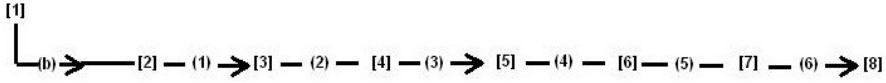
Figure 1: Pipeline Network, where: [x] vertices, (b) bombs, (x) tubes.

The pressurization and naphtha transportation in the pipeline are defined by specification of pressure (P) and external mass flow (W) directly in the vertices "1" and "8". The other parameters are calculated using fluid mechanics phenomenological equations.

To simulate the signal generated by real sensors, a time series containing random values was incorporated to the data generated by the simulator. The time series generated by the simulator present a 1-minute interval between each observation. Table 1 present all the sensors installed in the pipeline network.

Table 1: Sensors installed in the pipeline network

Sensor	Localization	Type	Sensor	Localization	Type
1	[1]	Flow	9	[7]	Press
2	[8]	Flow	10	[8]	Press
3	[1]	Press	11	(1)	Flow
4	[2]	Press	12	(2)	Flow
5	[3]	Press	13	(3)	Flow
6	[4]	Press	14	(4)	Flow
7	[5]	Press	15	(5)	Flow
8	[6]	Press	16	(6)	Flow

3. Results

This session describe the results obtained using the methodologies proposed here. Different approaches for detection, identification, localization and quantification of failures were tested. Several scenarios were studied, changing the location, severity and type of the tested failures. To simulate sensor failures, deviations in the form of step with magnitude 2%, 5%, 10% and 20%, were add to the signal from each sensor. Leaks caused by holes with a diameter of 1, 2, 4, 5, 8, 12, and 14 cm were simulated.

3.1. Sensor Failure Detection using hierarchical classification

A time series contain 450 observations of normal network operation and 50 observations of normal operation was employed (figure 2). Using Minkowski metric, more than 95% of the scenarios with sensors failures of severities 10% or 20% were correctly detected. The best performance was obtained using Minkowski metric (figure 3a). The degree of success of this methodology, however, is reduced in the presence of faults of low severity. The physical position of the sensor in the pipeline network also influences the efficiency of the detection tool. The fail detection in the sensors "3" and "10" presented lower efficiency.
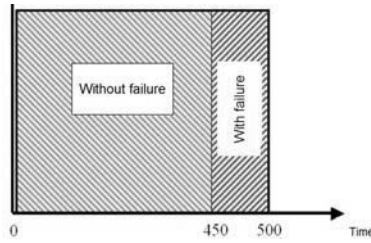
Figure 2: Phases with and without fault

| (a) | (b) |
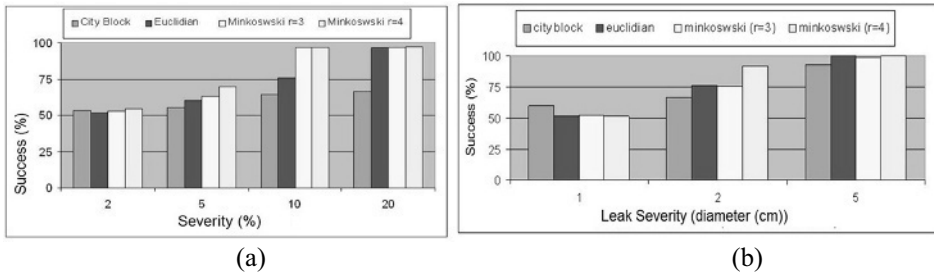
Figure 3: Degree of success to detect sensor failures (a) and leaks (b) using hierarchical classification

3.2. Leak Detection using hierarchical classification

Like sensor failure detection, hierarchical classification was employed to leak detection. Again, a time series consisting of 500 observations, of which 450 represent normal operating situation, and 50 represented a leak occurrence, was used. The metrics "Euclidian", "City Block", and Minkoswki" were tested.

Figure 3b presents the results for leaks with diameters of 1, 2 and 3 cm. The approach using hierarchical classification was not able to detect leaks of very small volume. For more severe leaks, the performance was satisfactory, especially using Minkoswki (r=4). The capability of the leak detector is also a function of the leak location.

3.3. Differentiating leaks from sensor failures occurrences

After an abnormal operation condition is detected, it is necessary to establish its type (sensor failure or leak). The method applied to detect failures is not able to differentiate between these two types of anomalies. A hybrid approach, using "hierarchical classification" and "method of prototypes" was applied. Twenty-two prototypes representing leaks and failures in sensors were created. The prototypes together with the data containing the detected anomaly are subjected to classification procedure. The similarity between the experimental data and the prototypes is able to identify the type of anomaly. This technique was successful in differentiating leaks from sensor failures, even in low severity situations (Table 2).

3.4. Localizing failures using prototypes

After determining the type of anomaly, the prototypes methodology is applied again in order to identify the location of the failure. The same 22 prototypes, 16 sensor failures and 6 cases of the "leak" are used. The most similar prototype is search using Sdist metric. This approach was effective to localize sensor faults and leaks in the pipeline network.

Table 2: Differentiating leaks from sensor failures occurrences

Leak (diameter - cm)	Success (%)	Sensor Failure (%)	Success (%)
1	33,33	1	100,00
2	83,33	2	100,00
4	100,00		
5	83,33		
8	83,33		
12	100,00		
14	100,00		

3.5. Quantification of sensor faults and leaks

Hierarchical classifications were applied to quantify sensor failures and leaks in pipeline network. The method can estimate the severity of operational deviation using the distance between the centers of the two clusters formed ("with" and "without" failure). Figure 4 shows the relation between the failure magnitude and distance between the two clusters. The Minkowski metric (r=4) has been used.
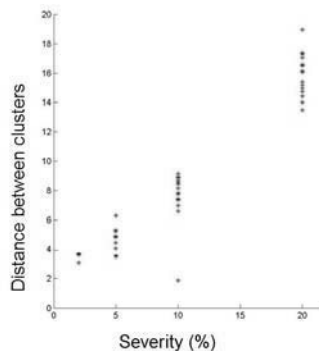
Figure 4: Distance between clusters vs. severity of the sensor failure

4. Conclusion

In this article, hierarchical classifiers were successfully used to detect abnormal operation in the pipeline network. The success of the methodology was strongly influenced by the severity of the failure and its localization. A hybrid approach, joining prototypes and hierarchical classifiers, was effective to identify the failure type (sensor failure or leak). Similar approach was applied to establish the location of the fault. Finally, the distance between the two clusters formed allows a quantification of the failure severity.

References

P. Hart, R. O. Duda, Pattern Classification, 2a ed. John Wiley Pro, 2000

A. Singhal, D. Seborg, 2002, Pattern Matching in Multivariate Time Series Databases Using a Moving-Window Approach, Ind. Eng. Chem. Res. v.41, p.3822-3838.

R. A. Silva, C. M. Buiatti, S. L. Cruz, J. A. F. R. Pereira, 1996, Pressure wave behaviour and leak detection in pipelines. Computers Chem. Engineering, v.20, p.S491-S496

W. Vandaele, 1983. Applied time series and Box-Jenkins Models. Academic Press

C. A. Vaz Junior, 2006, Detecção, Localização e Quantificação de Vazamentos: Uma Abordagem em Séries Temporais. Master in Science Thesis Escola de Química da UFRJ

Optimization of Wastewater Filtration Process in Submerged Membrane Bioreactors: Applicability of a Dynamic Model to Scale Up.

Alain Zarragoitia,[a, b] Silvie Schetrite,[a] Ulises J. Jáuregui-Haza,[c] Olivier Lorain,[d] Claire Albasi[a]

[a]*Laboratoire de Génie Chimique, UMR - CNRS 5503, 5 Rue Paulin Talabot, 31106 Toulouse cedex, France*
[b] *Departamento de Desarrollo Tecnológico, Centro de Química Farmacéutica (CQF), Calle 200 y 21, Apdo. 16042, Atabey, Playa. La Habana, Cuba*
[c]*Instituto Superior de Tecnologías y Ciencias Aplicadas (InSTEC), Ave. Salvador Allende y Luaces, Plaza, Ciudad de la Habana, Cuba*
[d]*POLYMEM, Toulouse cedex, France*

Abstract

The application of a hybrid mathematical model, which takes into account the effect of main variables in Submerged Membrane Bioreactor (SMBR) systems, was established in order to simulate and optimize the filtration process into SMBR, on the base of experimental data performed on a bench and pilot scale bioreactors. Numerical simulations of the wastewater treatment process were performed in order to find the optimal filtration conditions for both different scale bioreactors. Various operating conditions (idle-filtration time, aeration intensity, solids retention time, hydraulics retention time, and total suspended solids concentration) were tested. Different optimization criteria were considered to minimize the transmembrane pressure and energy consumption, and to maximize filtrate flow.

Keywords: Biological wastewater treatment; Submerged Membrane Bioreactor (SMBR); modeling; optimization.

1. Introduction

The wastewater treatment using a submerged membrane bioreactor (SMBR) allows obtaining a quite constant filtrate quality after biological degradation process and physical separation. The membrane bioreactor technology has been improved, and research on SMBR technology has increased significantly (Yang, 2006). MBR treatment of municipal wastewater yields high-quality water with reported removal percentages of 95%, 98%, and 99% (or greater) for chemical oxygen demand (COD), biochemical oxygen demand (BOD), and suspended solids (SS), respectively (Manem and Sanderson, 1996). Nevertheless, the filtration process is disturbed by the influence of complex factors that cause the membrane fouling.

On the other hand, mathematical modeling and simulation are powerful tools with which the specialists can predict the performances of potential systems under different operating conditions. For this reason, it is interesting to use models that can accurately describe the SMBR process for the design, and optimization of these systems.

The objective of this work is to show the practical application of a hybrid mathematical model, developed in a previous work (Zarragoitia, 2008), for optimizing the operation

of two SMBRs using different optimization criteria to minimize the transmembrane pressure (TMP) and energy consumption, and to maximize filtrate flow. The model was established considering Soluble Microbial Products (SMP) formation-degradation kinetic based on modified published models (Cho, 2003; Shane, 2007; Lu, 2001). A modification of Li and Wang's model (Li and Wang, 2006) allows to calculate the increase of the TMP evaluating the influence on fouling control of an intermittent aeration of coarse bubbles synchronized with the filtration cycles, and to analyze the effects of shear intensity on sludge cake removal. In order to describe the biological system behaviour a modified ASM1 model was developed to estimate the optimal operating conditions related to a specific bioreactor condition.

2. Experimental part

2.1. Experimental set-ups

Two different SMBRs were studied. In the first (bench scale), a U-shaped, hollow-fiber membrane (polysulfone, pore size: 0.1 μm, internal/external diameter: 0.4/0.7 mm) module with area of 0.3 m^2 (Polymem, Toulouse, France), was immersed in a bioreactor of 10.5 L of working volume. The municipal wastewater was continuously treated. The influent flow rate was controlled by the liquid level in the reactor. Filtration was operated in an intermittent sequence of filtration-relaxation. The TMP was continuously monitored (Sensor Keller). Filtrate flow was measured with an electromagnetic flow meter (Rosemount); a temperature sensor (PT 100; -50 to 250 °C) and a Mettler Toledo pH meter were used. The pH was maintained between 6.5 and 7.5. PC-based real time data acquisition hardware (IOTECK) and the software DASYLAB have been used for acquiring all data. The second bioreactor has four hollow-fibre membrane modules with an area of 2.5 m^2 and a bioreactor tank of 1 m^3 (Polymem, Labège, France). The membrane material is the same as the previous one. This bioreactor is operated in a largest plant, collecting wastewater from domestic and industrial areas. The bench scale bioreactor was operated with two types of aeration flows. An intermittent coarse bubbles flow injected closed to the fibres, to avoid the membrane fouling by reduction of cake formation, and a constant fine bubbles flow injected through a perforated membrane at the bottom of the reactor, providing mixing and biomass oxygenation. The membrane module was isolated to the contact with fine bubbles; thus, membrane fibers movement is only produced by the flow of coarse bubbles. The operation was stopped at 12 days of filtration or when the TMP reached 60 kPa under atmospheric pressure, and then a chemical cleaning was done.

2.2. Analytical methods

The Total Suspended Solids (TSS) concentration was determined using a PRECISA HA60 moisture analyzer. The activated sludge flocs size distribution was estimated with a laser granulometer (Mastersizer 2000, Malvern Instruments). The Extracellular Polymeric Substances (EPS) were quantified in influent and permeate samples. Proteins and polysaccharides were measured by spectrophotometric methods. The Lowry method was used for the quantification of proteins and humics. For quantitative analysis of carbohydrates, the anthrone method was used with glucose as standard. Respirometric measurements were used for the determination of readily and slowly biodegradable substrates considering $Y_H = 0.67$ gCOD / gCOD.

2.3. Simulation conditions and optimization procedure

A numerical simulation of the membrane fouling was performed using a program in Berkeley Madonna programming language. Membrane fouling development was

simulated by numerical iterations, which produces the overall TMP increase and the evolution in the distribution of the flux and the sludge cake layer across the membrane surface sections. The intermittent process of filtration and coarse bubbles injection were simulated by means of periodic functions. When the simulation results were compared with experimental data a good agreement was observed, thus, a D-Optimal Experimental Design (DOED) matrix was used for the optimization of the operating conditions for both SMBRs (Table 1). The values of the response variables in each experiment were estimated "in-silico" using our model. Finally, the predicted optimal operating conditions were validated by experimental way for bench scale SMBR. The validation of the optimum for the pilot scale SMBR is under study.

Table 1- Summary of the DOED.

	Bench scale SMBR, Experiments= 26, qa = 9 L/(m^2 s), J$_T$=0.24 m^3/(m^2 day)				
	tf (min)	tid (min)	Int$_{CB}$ (min)	t$_{CB}$ (min)	SRT (days)
Low actual	5	2	4	1	20
High actual	15	5	10	3	70
	Pilot scale SMBR, Experiments= 30, qa = 1 L/(m^2 s), J$_T$ = 0.24 m^3/(m^2 day)				
	tf (min)	tid (min)	Int$_{CB}$ (min)	t$_{CB}$ (min)	tbw (min)
Low actual	5	0.5	0.2	0.5	
High actual	105	1.5	0.75	3	0.5

qa= aeration intensity, tf= filtration time, tid= idle time, Int$_{CB}$= time interval between two coarse bubbles injection, t$_{CB}$= coarse bubbles injection time, SRT=solids retention time, J$_T$= overall flux tbw=backwashing time

3. Results and discussion

3.1. Optimization of the operating conditions for the bench scale SMBR.

During the optimization, different criteria were considered, to minimize the transmembrane pressure and energy consumption, and to maximize filtrate flow. The equation 1 shows the objective function used to calculate and to minimize the energy consumption. On the other hand, table 2 shows a summary of the measured influent characteristics and initial operating conditions used during simulation. The table 3 shows the optimization results estimated by simulation as well as the predicted values of the optimized response variables.

$$EC = \left[\frac{0.666(Qfil * TMP * tf)}{FV * (tf + tid)} + \frac{0.666(Qair * \Delta Pair * t_{CB})}{FV * (t_{CB} + Int_{CB})} \right] time \quad \langle \text{kW/m}^3 h \rangle \quad \text{(eq. 1)}$$

where: EC= Energy consumption (kW/m3h), FV= Filtrate volume (l), Qair= aeration flow (l/h), Qfilt= Filtration Flow (l/h), ΔPair= aeration drop pressure (bar)

During the optimization, only five factors were considered in the DOE. Figures 1-2 show different examples of the optimization results, but considering only the values of tid and tf in the axis. The other factors values could be calculated in the same way, once a specific optimization criterion is chosen. Table 4 points out the existence of differences in the tf and SRT values for the case A and B, but it also suggests that different filtration-idle time cycles may lead in both cases to work in optimal conditions, this also change the synchronization between the filtration and the intermittent aeration. This result is reasonable and may find explanation in the fact that the cases A and B have different influent characteristics and biomass concentration

(Table 2). These differences modify the behaviour of the system due to the presence of different F/M ratios, different sludge viscosities, and biological variables.

Table 2- Summary of the measured influent and initial conditions.

Case	Measured influent and initial conditions, values used during simulation
	X_{TSS0}	X_{TSSR0}	X_{S0}	X_{I0}	S_{S0}	S_{I0}	S_{SMP0}
A	25	6000	100	25	155	30	
B	30	8300	90	20	190	45	45

(Superscript 0 and R0 expresses "influent" and "initial value in the bioreactor" respectively), X_{TSS}= Total suspended solids in the mixed liquor (mg/l), X_S= slowly biodegradable products (mgCOD/l), X_I= inert organic materials (mgCOD/l), S_S= soluble biodegradable substrate (mgCOD/l), S_I= soluble inert matters (mgCOD/l), S_{SMP}=soluble microbial products (mgCOD/l)

Table 3- Optimization results calculated by simulation and the DOED for bench scale SBMR.

Case	\multicolumn{5}{c}{Optimized values of the operating variables}	\multicolumn{3}{c}{Predicted values of the response variables}
	tf	tid	Int_{CB}	t_{CB}	SRT	FV (12d)	TMP (12 d)	EC (12 d)
A	10 min	2 min	8 min	2 min	40 d	627 L	30 kPa	0.294 kW/m^{3}h
B	15 min	2 min	8 min	2 min	30 d	762 L	37 kPa	0.262 kW/m^{3}h
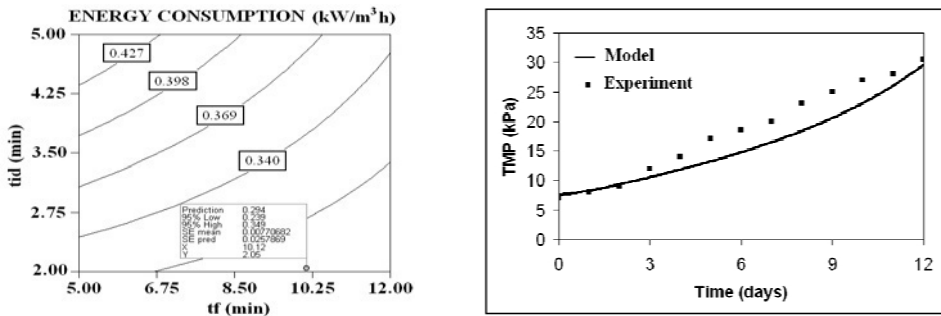
Figure 1- Optimization of operating conditions for the bench scale SMBR. Case A: Minimizing energy consumption [Left: Contour graph obtained during the optimization of the operating conditions (minimizing the energy consumption). Right: Experimental validation of the optimized values of the operating conditions].

The same observations can be made for the composition of mixed liquors, which are not the same for the cases A and B, and modifying the value of the specific filtration resistance of the sludge cake layer and, consequently, the fouling rate. The EC values were found between 0.15-0.96 kW/m^{3}h and the FV values between 432-800 l, therefore, the estimated optimal conditions facilitated, in both cases, to operate the bioreactor for an adequate volume of produced water with reasonable energy consumption (Table 3).

3.2. Optimization of the operating conditions for the pilot scale SMBR.

In the case of the pilot scale SMBR, the optimization study was carried out with the objective of modifying the usual values of the operating variables, in order to improve the operation efficiency of the SMBR. The results are shown in table 4.
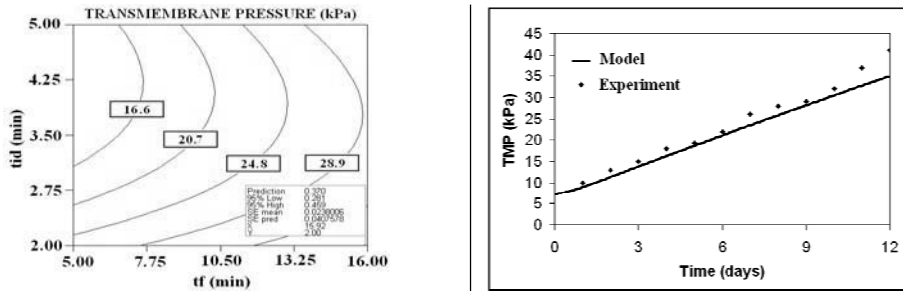
Figure 2- Optimization of operating conditions for the bench scale SMBR. Case B: Minimizing TMP [Left: Contour graph obtained during the optimization of the operating conditions (minimizing TMP). Right: Experimental validation of the optimized values of the operating conditions].

Table 4- Current and optimized values for the pilot scale SMBR (the values in brackets were used during the simulation with the optimized values of the operational variables).

Operating variables	Current value	Optimized value
tf (min)	5	7.61 (7.5)*
tid (min)	0.5	0.5
Int$_{CB}$ (min)	0.25	0.5
t$_{CB}$ (min)	0.25	1.76 (1.75)
tbw (min)	0.5	0.5

The analysis of the simulation-optimization results (Table 5) suggests that, for a better utilization of the SMBR, it would be convenient to increase the filtration time almost of 30% of its current value. In relation to the injection cycles of the coarse bubbles, the idle time between every injection could be increased in 80%, but at the same time, it is necessary to double the length of the aeration time. This operating way is more efficient because it decreases the energy consumption due to the aeration flow used to remove the sludge deposited on the membrane. Although, a higher idle time allows some temporary deposition of the sludge on the membrane, the simulation estimates that the increase of the injection time get possible to later efficiently the deposition of the sludge and the formation of the filtration cake on the membrane. On the other hand, a higher filtration time facilitates the production if a higher filtrate volume (increase of 11%) during the considered time, which decreases the estimated value of energy consumption by m^3 of treated water (decrease of 56%).

In relation to the value of the backwashing time, the model considers that the current value is correct (0.5 min). To evaluate the validity of the simulations and the quality of the model prediction, all results have still to be validated. Some simplifications have been made during the modeling of the system; therefore, it is necessary to characterize the influence of the assumptions and the limitations of the model on the optimization results. The optimization procedure presupposes the existence of a model calibrated to the real system, an accurate process of prediction by simulation is only possible after tuning the parameters of the model to a given plant. Therefore, the more sensible parameters must be recalibrated if the influent conditions and the biomass characteristics change considerably.
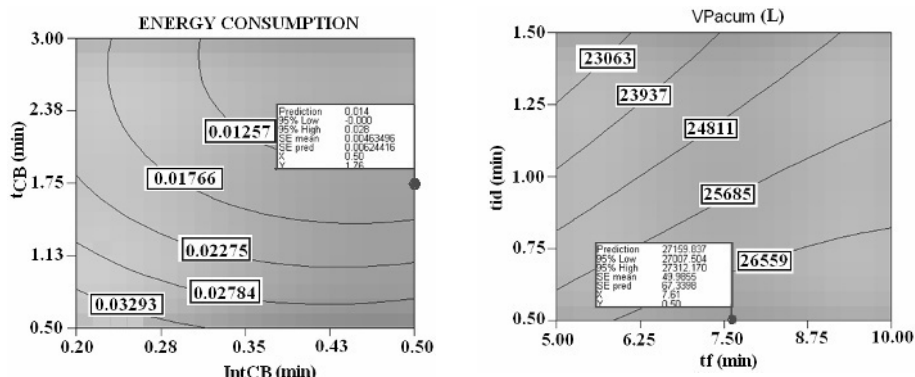
Figure 3- Examples of the contour graphs obtained during the optimization of the operating conditions for the pilot scale SMBR.

Table 5- Influence of the optimized operating variables on the response variables for the pilot scale SMBR.

	Reponse variables (at 15 days)		
	TMP (kPa)	VPacum (L)	EC (kW/hm^3)
Results with the current values of the operating variables	43.6	24000	0.055
Results with the optimized values of the operating variables	25	27160	0.024
TMP reduction	42.6 %	-	-
Volume increase	-	11.6 %	-
Energy consumption reduction	-	-	56.4 %

4. Conclusions

An optimization study to improve the filtration process on two different SMBRs at different scales has been successfully developed. A DOED was carried out "in silico" in order to optimize some operating variables for two different optimization criteria: minimizing energy consumption and minimizing TMP. Subsequently, the experimental verification process of the optimized operating conditions for the bench scale is in good agreement with the model predictions. Future tasks will focus on the system optimization, but, also including the aeration intensity and the filtration flux as factors to consider in the design of experiments. However, for better results a special attention is needed during parameters model calibration.

References

W. Yang, N. Cicek, J. Ilg, 2006, J. Membr. Sci., 270, 201–211.

J. Manem, R. Sanderson (Eds.), 1996, McGraw-Hill, New York.

A. Zarragoitia, S. Schetrite, M. Alliet, U. Jauregui-Haza, C. Albasi, 2008, J. Membr. Sci., 325, 2, 612–624.

J. Cho, K.-H. Ahn, Y. Seo, and Y. Lee. 2003,Wat. Sci. Tech., 47, 12, 177–181.

R. Shane, R. P. Merlo, W. Hermanowiczc, D. Jenkins, 2007, Water Res., 41, 947–958.

S.G. Lu, T. Imai, M. Ukita, M. Sekine, T. Higuchi, M. Fukagawa, 2001, Water Res., 35, 2038–2048.

X.-Y. Li, X.-M. Wang, 2006, J. Membr. Sci., 278, 151–161.

Control Structure Design for an Ethanol Production Plant

Gabriel V. N. de Andrade[a], Enrique L. Lima[b]

[a]*CHEMETCH - A Siemens Company, Rua da Quitanda, n. 50, 21° andar, Rio de Janeiro – RJ, CEP 20011-030, Brazil*
[b]*COPPE – Chemical Engineering Program – Federal University of Rio de Janeiro (UFRJ), Cidade Universitária, CP 68502, Rio de Janeiro – RJ, CEP 21945-970, Brazil*

Abstract

This paper studies two methodologies for control structure design. Both methodologies were applied to an ethanol production plant with an energy integration technology known as split-feed. A commercial process simulator was used. The proposed structures were tested to verify its performance. Finally, it was chosen the most efficient methodology for industrial applications.

Keywords: Plantwide control, ethanol, split-feed.

1. Introduction

Nowadays, there is great incentive for searching alternative energy sources to replace petroleum. Because of its renewable characteristics, ethanol has become an interesting option for automotive fuel. Also, industrial plants are now spending a great effort on reducing their energy costs. At ethanol production plants, a technology called split-feed has been used for this purpose. However, this technology still presents operating problems, demanding the development of control systems capable of dealing with such problems. A good alternative for this is the use of Plantwide control methodologies that, with detailed procedures, provide the necessary tools to build an efficient control structure.

The purpose of this paper is to apply plantwide control methodologies on an ethanol production plant and evaluate their practical use. Section 2 presents a short review of heat integration with split-feed and plantwide control. The ethanol production plant studied in this paper is presented in Section 3. In Sections 4 and 5, plantwide control methodologies are applied to the ethanol plant. A comparative analysis is made in Section 6 and the paper is concluded in Section 7.

2. The General Problem

2.1. Heat Integration with split-feed

The split-feed technology is used to thermally integrate distillation plants, where the feed is divided into two columns that operate at different pressures, as shown in Figure 1. The high pressure column condenser is responsible to generate the heat duty necessary to operate the low pressure column reboiler, reducing the consumption of heating and cooling fluids of the process.

Studies carried out by Lenhoff and Morari (1982), Chiang and Luyben (1988), Pohlmeier and Rix (1996) and Han and Park (1996) showed that several control

strategies have been tested in distillation systems with split-feed heat integration. Nevertheless, most of these control strategies were design for regular performance, dealing only with small process disturbances. The best results were achieved when a multivariable control scheme was used. Therefore, there is a need for development of more elaborated control structures that can handle the main problems of these complex distillation systems.
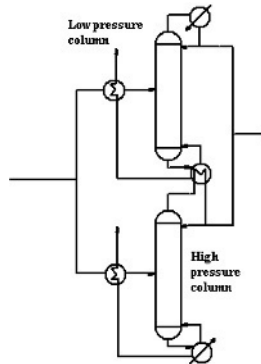
Figure 1: Split-feed arrangement

2.2. Plantwide Control

In the design of control systems the engineer must choose the best variables for measurement, manipulation and control. Also, it is very important to define how these variables should be interconnected, to guarantee the controllability of the plant as well as the good performance of the control system. Despite their great importance for design, these definitions and choices, which are part of the structural decisions (Skogestad, 2000), are usually made based on the engineer experience on the process.

Plantwide control is the design of the control system of an overall plant, where the engineer defines the control philosophy with emphasis on the structural decisions. Due to the complexity of the plantwide control problem, it is very difficult to use a mathematical approach, especially with respect to the dynamic and steady state modeling of the plant.

McAvoy and Ye (1994), Price et al. (1994), Stephanopoulos and Ng (2000) and Skogestad (2004) presented several methodologies for control structure design using the plantwide control concept. However, some of these methodologies use optimization tools that demand more complex computational approaches. Also, due to the lack of use of these methodologies in the design of industrial plants (maybe because the engineers still don't trust them), it is important to change this scenario by implementing them in different kinds of plants.

3. The ethanol production plant

In this paper, control structures will be proposed to an existing ethanol production plant (distillation and stripping sections), which is described below.

The unit has two sections, each one operating at a different pressure. The distillation section has a split-feed arrangement. With this heat integration, the plant has significant reduction on the utilities consumption (steam and cooling water), but the strong interaction between the two sections causes several problems to the control system.

Due to the great amount of streams and equipment in the plant, there are around 80 possible measurements and 50 manipulated variables, which give lots of possibilities of pairing. Therefore, in this paper, only the low pressure section will be considered, changing the scenario to 30 possible measurements and 15 manipulated variables. Figure 2 presents the low pressure section of the unit:
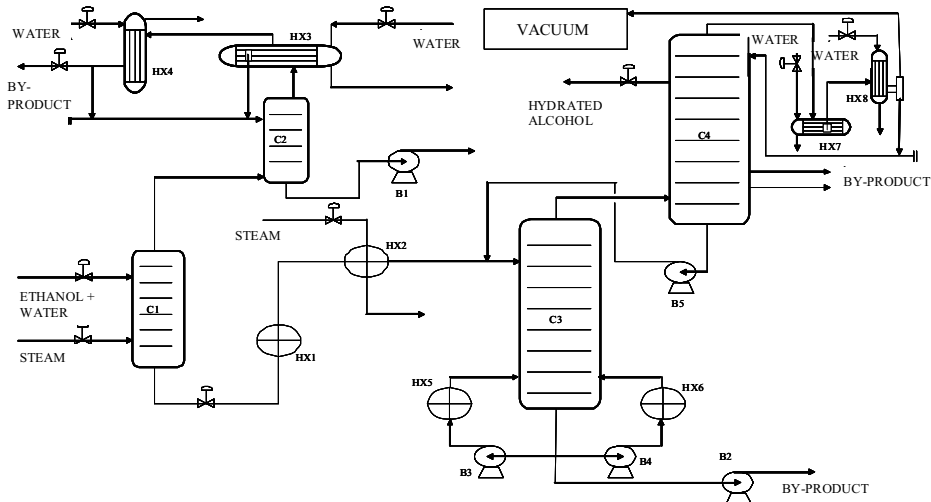
Figure 2: Low pressure section.

3.1. Simulation

The plant was simulated in a commercial process simulator software, ChemCad 5.6.4 (ChemStations). For the control loops, the PID module of ChemCad (traditional feedback controller) was used.

Since this section of the plant operates at a very low pressure, there is a specific vacuum system to control it. This system could not be simulated in ChemCad, so the pressure was kept constant in the simulation, even during the dynamic tests.

Looking for methodologies that can be easily understood and applied in the industry, the procedures proposed by McAvoy and Ye (1994) and Price et al. (1994) were chosen to be used in this work. This choice was based firstly on the fact that none of these methodologies necessarily uses optimization tools and/or multivariable control. The use of these tools would demand a lot of compilation and execution hours, as well as a computational structure capable of dealing with complex problems. In addition, both procedures are partially based on the engineers` knowledge of the process and also on analysis tools that can be easily applied.

4. Application of the methodology proposed by McAvoy and Ye (1994)

McAvoy and Ye (1994) proposed a systematic methodology for control structure design based on a Plantwide control approach, where single-input-single-output (SISO) control loops are used. The methodology is divided into four stages, as described below:

At the first stage, the fast response control loops are defined. Usually, these loops are simple flows and utility temperatures. The second stage has four steps: closing the level loops; interaction, stability and saturation analysis; steady state disturbance analysis and; tuning and testing via dynamic simulation. In the third stage, the quality control loops are defined, through a global material balance. The last stage consists of adding upper control layers, using real-time optimization tools, predictive control and others.

Besides the last stage, all the others were applied to the ethanol plant using the methodology of McAvoy and Ye (1994). The loops listed in Figure 4 were tuned and the control system was tested dynamically in order to verify its performance. The tests consisted of inserting disturbances (steps) on the unit feed flow and thermal conditions. The control system showed good results, rejecting satisfactorily the external disturbances, as can be seen in Figure 3:
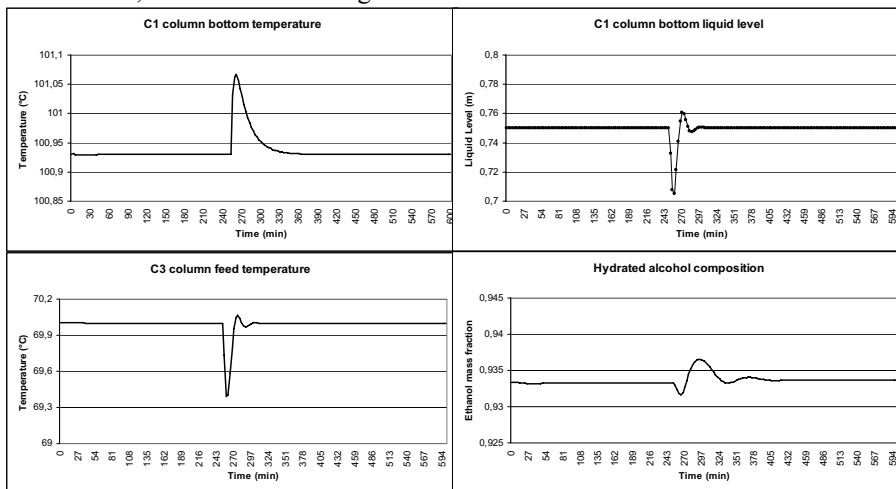
Figure 3: Tests with the control system proposed using the methodology of McAvoy e Ye (1994)

5. Application of the methodology proposed by Price et al. (1994)

Price et al. (1994) proposed a methodology for control structure design based mainly on the unit inventory management and the production rate.

The procedure starts by identifying the primary path of the process and choosing a stream to control the production rate. Then, a set of inventory control loops is added, making sure that all changes in the production rate will be propagated throughout the process.

According to the authors, the production rate manipulated variable must be chosen to obtain a self-consistent control chain, which is defined as a control chain capable of transmitting a production rate change throughout the plant, making adjustments in all inlet and outlet streams. When the control chain is not self-consistent, it is necessary to add extra control loops to help propagating production rate changes.

All the steps of the methodology of Price et al. (1994) were applied, resulting in three different control systems for the plant. As recommended by the authors, the control system with an internal flow controlling the production rate was chosen. The control loops, presented in Figure 4, were then tuned and tested via dynamic simulation, with the same disturbances used at the control structure proposed with the methodology of McAvoy and Ye (1994). Some of the obtained results can be seen in Figure 5.

Manipulated variables	McAvoy and Ye (1994) Methodology	Price et al. (1994) Methodology
	Controlled variables	
C1 inlet stream flow valve	C1 feed flow	C1 bottom liquid level
C4 side draw flow valve	C4 side draw flow	Production rate/C4 side draw flow ratio
C2 distillate flow valve	C2 distillate flow	C2 distillate flow
HX3 cooling water flow	HX3 cooling water temperature	-
HX7 cooling water flow	HX7 cooling water temperature	C4 top temperature
HX8 cooling water flow	HX8 cooling water temperature	-
C3 bottom stream flow	C3 bottom liquid level	C3 bottom liquid level
C4 bottom stream flow	C4 bottom liquid level	C4 bottom liquid level
C2 bottom stream flow	C2 bottom liquid level	C2 bottom liquid level
C1 bottom stream flow	C1 bottom liquid level	Production rate
Reboiler flow	C3 bottom temperature	C3 bottom temperature
Direct steam flow	C1 bottom temperature	C1 bottom temperature
HX2 steam flow	C3 inlet stream temperature	C3 inlet stream temperature
HX4 cooling water flow	C2 distillate temperature	C2 distillate temperature
HX7 cooling water temperature (setpoint)	C4 top temperature	-
C4 side draw flow (setpoint)	C4 bottom temperature	-

Figure 4: Control loops defined for each structure.
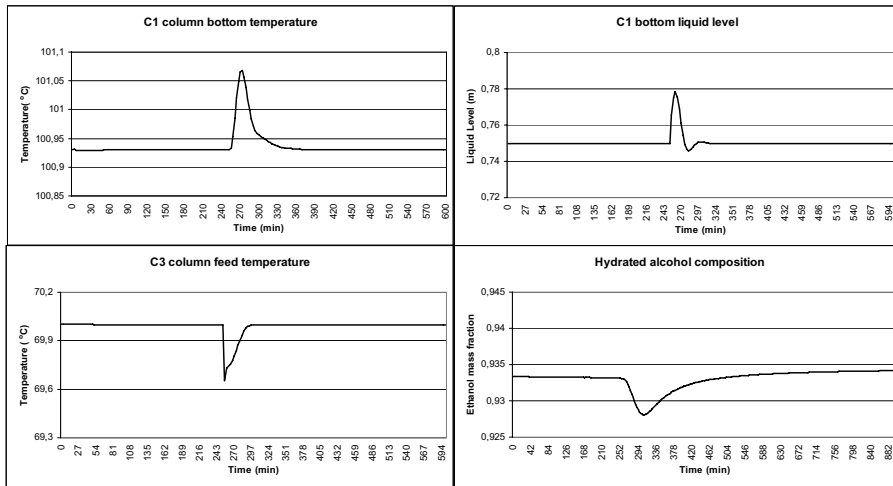
Figure 5: Tests with the control system proposed using the methodology of Price et al. (1994).

6. Comparative Analysis

The methodology proposed by McAvoy and Ye (1994) shows that it is possible to design a simple control structure in a fast and practical way, by using some analysis tools combined with the experience of the engineer on the process.

The methodology proposed by Price et al. (1994) presents a procedure that is easy to follow and understand. The methodology is based on the production rate and inventory controls, where the last ones are responsible for transmitting changes in the former throughout the process. McAvoy and Ye (1994) also give a significant importance to inventory control, dedicating one stage to determine level control loops. However, the production rate control is more detailed by Price et al. (1994).

Another difference between the methodologies is on the procedure to define the control loops that are not related to inventory and production rate control. McAvoy and Ye (1994) use interaction, stability and saturation analysis tools to determine the decentralized control loops. On the other hand, Price et al. (1994) complete their control structure by only defining product quality control loops.

With respect to production rate control, the methodology of Price et al. (1994), with the self-consistent concept, assures that changes in the production rate are totally propagated through the whole plant. Concerning this matter, McAvoy and Ye (1994) don't give any guideline to perform this kind of analysis.

Both methodologies strongly depend on the knowledge and experience of the engineer on the process. Therefore, it is very important to use some theoretical tools to check stability, saturation, interaction and any other problem that can occur with the control system. Even though both methodologies have a great applicability, the one proposed by McAvoy and Ye (1994) seems to be the most adequate for industrial applications, since it uses important analysis tools combined with process knowledge.

Also, the use of the self-consistent concept and the production rate manipulation analysis by Price et al. (1994) could be adapted to the methodology proposed by McAvoy and Ye (1994), generating a more complete procedure for control structure design.

7. Conclusions

The plantwide control literature presents several methodologies and procedures for control structure design. These methodologies, in most cases, have a process oriented approach, but some of them also use analysis and optimization tools.

The methodologies proposed by McAvoy and Ye (1994) and Price et al. (1994) were chosen to be studied on an ethanol production plant. For the first one, the stages were applied and the obtained control structure was tested to verify its performance in the presence of external disturbances. The control structure presented good results. For the methodology by Price et al. (1994), the given guidelines were followed to obtain a final control structure, which was also tested, showing good responses against external disturbances. Finally, a comparative analysis was made and the methodology by McAvoy and Ye (1994) was chosen the best for this industrial application.

References

Chemstations, CHEMCAD Simulation Suite.

T.P. Chiang, W.L. Luyben, 1988, Comparisons of the dynamic performances of three heat integrated distillation configurations, Ind. Eng. Chem. Res., v. 27, pp.99-104.

M. Han, S. Park, 1996, Multivariable Control of Double-effect Distillation Configurations, Journal of Process Control v. 6, n. 4, pp. 247-253.

A.M. Lenhoff, M. Morari, 1982, Design of resilient processing plants - I. Process design under consideration of dynamic aspects, Chemical Engineering Science, v. 37, pp.245-258.

T.J. McAvoy, N. Ye, 1994, Base Control for the Tennessee Eastman Problem, Computers Chem. Engng., v. 18, n. 5, pp.383-413.

J. Pohlmeier, A. Rix, 1996, Interactive Plant and Control Design of a Double-Effect Distillation Column, Computers and Chemical Engineering, v. 20, n. 4, pp. 395-400.

R.M. Price, P.R. Lyman, C. Georgakis, 1994, Throughput Manipulation in Plantwide Controle Structures, Ind. Eng. Chem. Res., v.33, pp. 1197-1207.

S. Skogestad, 2004, Control structure design for complete chemical plants, Computers and Chemical Engineering, v. 28, pp. 219-234.

G. Stephanopoulos, C. Ng, 2000, Perspectives on the synthesis of plant-wide control structures, Journal of Process Control, v.10, pp. 97-111.

Nonlinear Dynamic Process Monitoring Using Canonical Variate Analysis and Kernel Density Estimations

P.P. Odiowei[a] and Y. Cao[a]

[a]*School of Engineering, Cranfield University, Bedford, MK43 0AL, UK*

Abstract

Amongst process monitoring techniques the Principal Component Analysis (PCA) and the Partial Least Squares Regression Analysis (PLS) assume that the observations at different times are independent. However, for most industrial processes, these assumptions are invalid because of their dynamic features. For dynamic processes, the Canonical Variate Analysis (CVA) based approach is more appropriate than the PCA and the PLS based approaches. The CVA model is linear and control limits associated with the CVA are traditionally derived based on the Gaussian assumption. However, most industrial processes are non-linear and the Gaussian assumption is invalid for such processes so that techniques based on this assumption may not be able to correctly identify underline faults. In this work, a new monitoring technique using the CVA with control limits derived from the estimated probability density function through kernel density estimation (KDE) is proposed and applied to the Tennessee Eastman Process Plant. The proposed CVA with KDE approach is able to significantly improve the monitoring performance compared to other methods mentioned above.

Keywords: Canonical Variate Analysis, Probability Density Function, Kernel Density Estimation, Process Monitoring

1. Introduction

Process monitoring is essential to maintain high quality products and process safety. Process monitoring techniques like the PCA and the PLS rely on static models and assume that the observations are independent in time and follow a Gaussian distribution. However, such assumptions are invalid for most chemical processes because measurements driven by noise and disturbances are strongly auto-correlated. Therefore the static PCA and PLS based approaches are inappropriate to monitor such dynamic processes. To address this issue, Ku et al.[1] proposed a dynamic extension of PCA (DPCA) by using lagged variables to develop dynamic models through parallel analysis. However, diagnosis of abnormal behaviour becomes more complicated with DPCA and the principal components extracted in this way are not necessarily the minimal dynamic representations[2].

More recently, monitoring techniques based on Canonical Variate Analysis (CVA) have been developed with control limits derived based on the Gaussian assumption[3,4,5]. CVA is a monitoring technique that is based on state variables which are linear combinations of the past measurements. Norvalis et al.[5] developed a process monitoring tool that combined CVA and knowledge based systems (KBS) with control limits of the T^2 metric. Juan and Fei[4] employed CVA for fault detection based on T^2 charts while

Chiang et al.[3] employed CVA to include both input and output variables to estimate state variables with control limits of the T^2 and Q metrics to measure the state and residual spaces.

Traditionally, control limits of the T^2 metric for the state space and noise space and the Q metrics are estimated based on an assumption that the latent or state variables follow a normal distribution. However, for most industrial processes, these assumptions are invalid because of their dynamic and non-linear properties. In such a case, the control limit estimated based on the Gaussian assumption is unable to correctly identify underline faults. The problem of monitoring non-Gaussian processes can be addressed by estimating the underline probability density function (PDF) of the T^2 and Q metrics through the kernel density estimation (KDE) to derive correct control limit[6,7,8]. Martin and Morris[6] presented an overview of the PCA and the PLS with control limits on T^2 and M^2 metrics. The M^2 metric was estimated based on the PDF and was more efficient than the T^2 metric. Chen et al.[7] adopted the PCA with KDE approach to monitor a gas smelter process and demonstrated that the KDE technique could be used as a tool to obtain nonparametric empirical density function for a more efficient process monitoring. Xiong et al.[8] presented a study of Independent Component Analysis with KDE to monitor a catalyzer reactor.

In this paper, control limits of the T^2 and Q metrics associated with CVA are estimated based on kernel density estimations resulting in a new extension of the CVA algorithm, the 'CVA with KDE' for nonlinear dynamic process monitoring. For comparison, PCA and PLS with and without KDE as well as CVA with and without KDE are considered in the present study. The monitoring performance is significantly improved by using the 'CVA with KDE' compared with other monitoring algorithms mentioned above. This paper is organised as follows: Section 2 explains the CVA model while section 3 describes monitoring metrics and their control limit derived through Kernel Density Estimations. The procedure of CVA with KDE is then summarised in section 4 and the application of the proposed CVA with KDE to the Tennessee Eastman Process plant described in section 5. Finally, the work is concluded in section 6. Up to authors knowledge, this is the first time the KDE approach has been applied to estimate control limits for CVA based monitoring techniques. This establishes the main contribution of this work.

2. Canonical Variate Analysis (CVA)

Canonical Variate Analysis (CVA) is a dimension reduction technique based on a linear dynamic state space model to estimate state variables for dynamic process monitoring. Assume $y(k) \in R^m$, for k=1, 2,…,N are process data collected under a normal operation condition. Variables that occur before time k, are described as the past variables ($y_p(k)$) while variables that occur after time k, are the future variables ($y_f(k)$). More specifically, the past and future output vectors are determined in (1) and (2), respectively. Based on these definitions, the past and future Hankel output matrices are obtained in (3) and (4), respectively.

$$y_p(k) = [y^T(k\text{-}1), y^T(k\text{-}2), y^T(k\text{-}3),…. y^T(k\text{-}p)]^T \in R^{mp} \qquad (1)$$
$$y_f(k) = [y^T(k), y^T(k\text{+}1), y^T(k\text{+}2),…. y^T(k\text{+}f\text{-}1)]^T \in R^{mf} \qquad (2)$$
$$Y_p = [y_p(k\text{-}1)\ \ y_p(k\text{-}2)\ \ y_p(k\text{-}3)… y_p(k\text{-}M)] \in R^{mp \times M} \qquad (3)$$
$$Y_f = [y_f(k)\ \ y_f(k\text{+}1)\ \ y_f(k\text{+}2)…. y_f(k\text{+}M\text{-}1)] \in R^{mf \times M} \qquad (4)$$

where mp and mf are the lengths of the past and future output vectors respectively and M is the number of lags. CVA maximises the correlation between the past and future variables to explain the information in the process data. Precisely, the covariance matrices of the past and future Hankel matrices are $\sum_{pp} = E(Y_p Y_p^T)$ and $\sum_{ff} = E(Y_f Y_f^T)$ respectively, and the cross covariance between the future and the past Hankel matrices is $\sum_{fp} = E(Y_f Y_p^T)$. However, the realisation of CVA requires the covariance of the past and future and the cross covariance between past and future Hankel matrices to be conditioned against any singularities by taking their square roots[2,4]. From the square root of the covariance matrices, a scaled Hankel matrix (H) is formed and the singular value decomposition (SVD) of H is shown in (5).

$$H = \sum_{ff}^{-1/2} \sum_{fp} \sum_{pp}^{-1/2} = U \sum V^T \tag{5}$$

The full singular value (SV) matrix $\sum \in R^{mf \times mp}$ contains the SVs in the diagonal of $\sum$ in a descending order. $U \in R^{mf \times mf}$ contains the left singular-vectors and $V \in R^{mp \times mp}$ contains the right singular-vectors. The ratio of the specific SV to the sum of all the SVs is employed to determine n, which is the order of the system[4,9]. The state variables are linear combinations of the transformation matrix (J) which is given in as;

$$J = U_n^T \sum_{pp}^{-1/2} \tag{6}$$

In this way, the state variables estimated in (7) have identity covariance matrix, **I**.

$$x_k = U_n^T \sum_{pp}^{(-1/2)} y_p(k) = Jy_p(k) \tag{7}$$

Hence the state variable will have n linear combinations of the past in the state space and (mp-n) linear combinations in the noise space.

3. Control Limit through Kernel Density Estimations

For monitoring, the T^2 and Q metrics are applied to the states derived in (7) to monitor the state and residual spaces respectively. The T2 metric is defined in (8)

$$T^2(k) = [x_k - \bar{x}]^T S^{-1}[x_k - \bar{x}] \tag{8}$$

for the state space as well and the noise space, where $\bar{x}$ is the mean of the sample from which the CVA model is developed, while S is the covariance and is estimated in (9).

$$S = \frac{1}{N-1} \sum_{k=1}^N [x_k - \bar{x}][x_k - \bar{x}]^T \tag{9}$$

N is the number of samples. The residual vector (e) is defined as;

$$e = (I - J^T J)y_p(k) \tag{10}$$

From (10) above, the Q-statistic is estimated as in (11) below.

$$Q = e^T e \tag{11}$$

One solution of monitoring non-Gaussian processes is to estimate the PDF directly for the T^2 and Q metrics by a non-parametric approach such as the KDE[6,7,8,10]. Assume x is a random variable and the density function is denoted by p(x). This means that

$$P(a < x < b) = \int_a^b P(x)dx \tag{12}$$

Therefore, an appropriate control limit can be derived for a specific confidence bound from (12) if p(x) is known. To estimate the PDF p(x) at point x, the kernel density estimator $\hat{p}_h(x)$ is defined in (13) based on the assumption that Xi is a set of random samples with size N, K is a kernel function and h is the bandwith.

$$\hat{p}_h(x) = \frac{1}{Nh} \sum_{i=1}^{N} K(\frac{x - X_i}{h}) \tag{13}$$

Then $\hat{p}_h(x)$ determined in (13) is the estimated PDF and based on the density estimation, appropriate control limits corresponding to the given confidence level is determined to detect the underline faults in the non-Gaussian process data.

4. CVA with KDE

The procedure of the proposed CVA with KDE algorithm consisting of the CVA modelling, the KDE estimation and the monitoring stage is shown below;

Building the CVA model from the normal operating process
1.   Collect data from normal operation for CVA
2.   Determine past and future output vectors as in (1) and (2)
3.   Determine the Hankel output matrices as in (3) and (4)
4.   Estimate the covariance of the past, future and cross covariance between past
          and future hankel matrices as demonstrated in section 2 above
5.   Estimate the scaled hankel matrix and perform SVD as in (5)
6.   Estimate the canonical variates as in (6)
7.   Estimate the state variables as in (7)

KDE Estimation
8.   Estimate the T^2 and Q metrics as in (8) and (11)
9.   Estimate the PDF of the T^2 and Q metrics using KDE in (13)
10.   Determine the control limits corresponding to a given confidence interval
          based on (12).

Monitoring Stage
11.   Collect fresh data for monitoring
12.   Calculate current states and residual using fresh measurements
13.   Calculate the T^2 and Q metric of current states and residual
14.   Employ the control limits estimated from step 10 for process monitoring

5. Application- Tennessee Eastman Process Plant

The Tennessee Eastman Process (TEP) plant[11] has 5 main units which are the reactor, condenser, separator, stripper and compressor[3]. The TEP data consists of two blocks; the training and test data sets each of which has 52 measured variables and 22 scenarios corresponding to Fault 0 to Fault 21, with Fault 0 being the data simulated at normal operating condition (no fault) and Faults 1 – 21 corresponding to data sets from the simulated fault processes, each with a specified fault. For the purpose of this study, six TEP faults have been employed which are Faults 2, 6, 10, 16, 19 and 21. All the results in this study have been based on a 99% confidence level. The superiority of the proposed CVA with KDE is compared with other monitoring techniques using the Q metric as shown in Figure 1 while Figure 2 shows the advantage of the new method being able to detect faults earlier than other monitoring techniques.
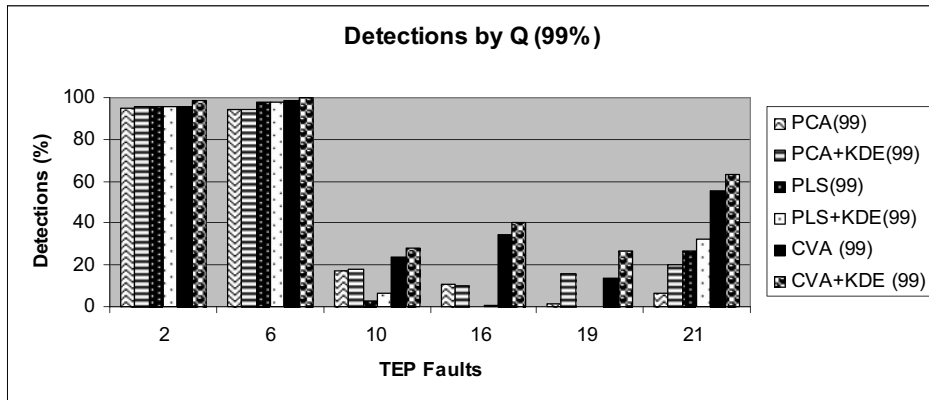
Figure 1 Comparison of All Monitoring Techniques (Q-99%)

For all six faults considered, Figure 1 clearly shows that the proposed CVA with KDE is able to significantly improve the monitoring performance compared with other monitoring techniques employed in this study, i.e. PCA, PCA with KDE, PLS, PLS with KDE and CVA only.
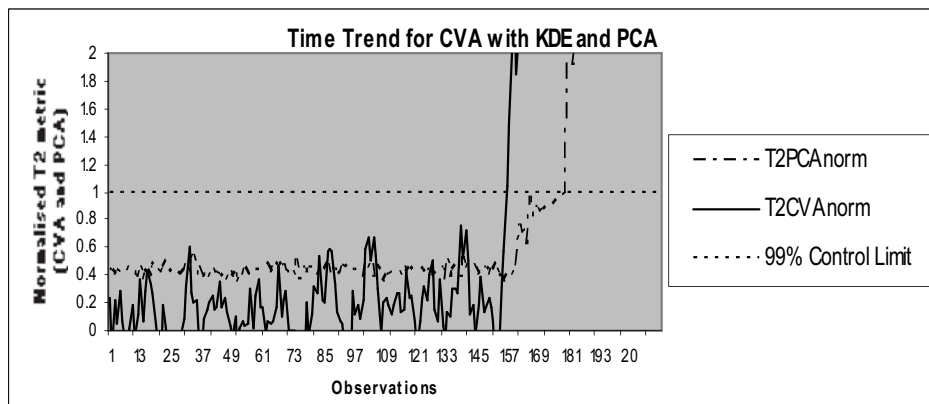
Figure 2 Time Trend for Fault 6 based on PCA and CVA with KDE

In addition to improving the performance significantly, the CVA with KDE approach is also able to detect faults earlier than other monitoring approaches so that operators have more time to deal with the underline fault before any disaster happens. For fault 6, Figure 2, shows two normalised T^2 metric curves generalised by the PCA and the CVA with KDE respectively. It shows that the PCA detects the fault at the 183rd observation while the CVA with KDE detects the fault at the 160th observation, which is about sixty nine minutes earlier than the PCA with a much more clear indication for a fault.

6. Conclusions

In this study a CVA with KDE approach is proposed for non-linear dynamic process monitoring, where due to the nonlinearity, the normal Gaussian assumption on measurements is invalid. This is the main contribution of this work. The Tennessee Eastman Process Simulation data was employed as the case study. The monitoring

performance of the proposed CVA with KDE was compared with PCA and PLS with and without KDE as well as CVA without KDE. The time trends were also investigated for the CVA with KDE compared with the time trend for PCA. The comparison demonstrates that the proposed CVA with KDE approach is able to improve the monitoring performance over other monitoring techniques mentioned above. Furthermore, the CVA with KDE is found to be able to detect faults earlier than the PCA. In summary, the CVA with KDE as a tool for nonlinear dynamic process monitoring is more appropriate and more efficient than the traditional CVA and other approaches mentioned above.

References

1. W. F. Ku, H. R. Storer and C. Georgakis (1995). Disturbance Detection and Isolation by Dynamic Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems, Pages 179 – 196.

2. A. Negiz and A. Cinar. Monitoring of Multivariable Dynamic Processes and Sensor Auditing. Journal of Process Control, Vol. 8, No. 56, 1998, Pages 375 – 380.

3. L. H Chiang, E. L. Russell and R. D. Braatz (2001). Fault Detection and Diagnosis in Industrial Systems. Springer, London, Pages 85-98, 103-109.

4. L. Juan and L. Fei. Statistical Modelling of Dynamic Multivariate Process using Canonical Variate Analysis. Proceedings of IEEE International Conference of Information and Automation, Colombo, Sri Lanka, 15 -17, December 2006, Pages 218-221.

5. A. Norvalis, A. Negiz, J. DeCicco and A. Cinar (2000). Intelligent Process Monitoring by Interfacing Knowledge-Based Systems and Multivariate Statistical Monitoring. Journal of Process Control, Vol. 10, Pages 341 – 350.

6. E. B. Martin and A. J. Morris (1996). Non-Parametric Confidence Bounds for Process Performance Monitoring Charts. Journal of Process Control, Vol. 6, No. 6, Pages 349-358.

7. Q. Chen, P. Goulding, D. Sandoz and R. Wyne. Application of Kernel Density Estimates to Condition Monitoring for Process Industries. Proceedings of the American Control Conference, Philadelphia, PA, USA, 21-26 June 1998.

8. L. Xiong, J. Liang and J. Qian, J (2007). Multivariate Statistical Process Monitoring of an Industrial Polypropylene Catalyzer Reactor with Component Analysis and Kernel Density Estimation. Chinese Journal of Chemical Engineering, Vol. 15, No. 4, Pages 524 – 32.

A. Negiz and A. Cinar (1997). PLS, Balanced and Canonical Variate Realization Techniques for Identifying VARMA models in State Space. Chemometrics and Intelligent Laboratory Systems, Vol. 38, Pages 209 – 221.

T. Chen, J. Morris and E. Martin (2006). Probability Density Estimation via an Infinite Gaussian Mixture Model: Application to Statistical Process Monitoring. Appl. Statist, Vol. 55, Part 5, 2006, Pages 699 – 715.

N. L. Ricker (2001). Tennessee Eastman Challenge archive. Available at: http://depts.washington.edu/control/Larry/TE/download.html.

1563

Multi-Period Continuous-Time Formulation for Integrated Scheduling, Blending, and Distribution of Refinery Products

Jie Li,[a] I. A. Karimi,[a] Rajagopalan Srinivasan[a,b]

[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576*
[b]*Process Sciences and Modeling, Institute of Chemical and Engineering Sciences, 1 Pesek Road, Jurong Island, Singapore 627833*

Abstract

In this paper, we develop a slot-based multi-period mixed integer linear programming (MILP) formulation for an integrated treatment of recipe, specifications, blending, storage, and distribution, and incorporate many real-life features such as multi-purpose product tanks, parallel non-identical blenders, minimum run lengths, changeovers, etc. To enforce constant rates during blending runs, we develop a novel and efficient procedure that solves successive MILPs instead of a non-convex MINLP. We use fourteen examples of varying sizes and features to illustrate the superiority and effectiveness of our formulation and solution approach.

Keywords: Gasoline, recipe, blending and distribution, non-convex mixed-integer nonlinear programming (MINLP), property index

1. Introduction

The overall refinery operations (Pinto et al. 2000) involve crude oil storage and processing, intermediate processing, and product blending and distribution. Scheduling of crude oil operations (Reddy et al. 2004a,b; Li et al. 2007) has received the most attention so far. However, only limited work exists on the scheduling of product blending and distribution operations.

Gasoline is one of the most profitable products of a refinery and can account for as much as 60-70% of total profit. However, this process involves nonlinear blending and complex combinatorics, and can easily result in suboptimal schedules and costly quality give-aways. The large numbers of orders, delivery dates, blenders, blend components, tanks, quality specifications, etc. make this problem highly complex and nonlinear.

So far, several works in the literature (Pinto et al., 2000; Glismann & Gruhn, 2001; Mendez et al., 2006) have addressed the problem of product blending operations and incorporated some real-life features such as variable recipes, identical parallel blenders, etc. However, these works did not integrate the distribution operations with blending, and did not force the blending rate to be constant in a run and minimum run length. Jia and Ierapetritou (2003) proposed an MILP model for scheduling gasoline blending and distribution operations simultaneously. However, their model lacked many key operation features such as multiple parallel non-identical blenders, variable recipes, etc.

In this paper, we develop a multi-period MILP formulation that incorporates several real-life operation features such as multi-purpose product tanks, parallel non-identical blenders, minimum run lengths, changeovers, linear property indices (Li et al. 2007), piecewise constant profiles for blend component qualities and feed rates, etc. To ensure

constant rates in blending runs, we develop a novel schedule adjustment procedure that solves only MILPs and no non-convex MINLP.

2. Problem Statement

Consider a gasoline blending and distribution unit (GBDU) in a typical refinery (Figure 1). It employs I component tanks ($i = 1, 2, \ldots, I$), B blenders ($b = 1, 2, \ldots, B$), J product tanks ($j = 1, 2, \ldots, J$), and some lifting ports. Each component has a distinct, and known quality or specification, and component i is stored in its own dedicated component tank i. At time zero, it has O orders ($o = 1, 2, \ldots, O$) to fulfill during the coming scheduling horizon $[0, H]$. Each order o involving a single product has a time window $[DD_o^L, DD_o^U]$ for delivery. Any delivery after DD_o^U incurs a demurrage cost (DM_o). The quality of blend components is specified in terms of various property indices such as RBN (Research Octane Number Index). With this, the GBDU problem addressed in this paper can be stated as:
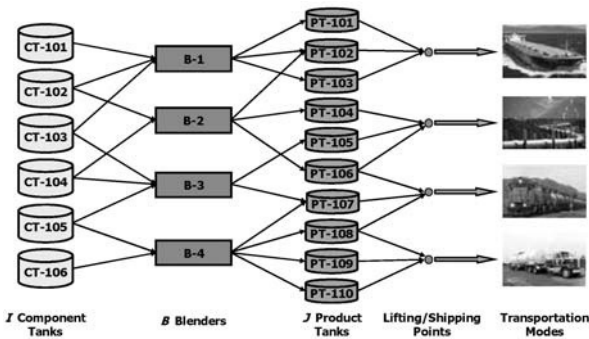
Figure 1 Schematic of gasoline blending and distribution

Given:
1. I components, their property indices and P products, their property indices limits.
2. I component tanks, initial inventories, capacity limits, feed flow profiles into tanks.
4. B blenders, allowable products, minimum blending length, and blending rates limits.
5. J Product tanks, allowable products stored in each tank, capacity limits, initial products and holdups, and delivery (lifting) rates for various products.
6. O orders, their constituent products, amounts, and delivery time windows.
Determine:
1. The blenders that each component tank should feed over time, and at what flow rates.
2. The products that each blender should produce over time, and at what rates.
3. The products that each product tank should receive over time, from which blender, and at what flow rates.
4. The orders that each product tank should deliver over time and their amounts.
Assuming:
1. Flow rate profile of each component from the upstream process is piecewise constant.
2. Component quality profile is also piecewise constant.
Other assumptions and some operating rules are not presented here.

3. Multi-Period MILP

Given the rate profiles of feeds into component tanks and profiles of component qualities, we divide the horizon H into T periods ($t = 1, 2, \ldots, T$) of lengths H_t such that

the flow rates of components and component qualities are constant in each period and $H = H_1 + H_2 + \ldots + H_T$. We follow the approach used by Karimi and McDonald (1997) in their second model (M2). Thus, we divide each period into several process-slots (Liu and Karimi, 2007) of variable lengths. The process-slots are common to or synchronized across all units (tanks and blenders). Let K (k = 1, 2, …, K) be the total number of slots and $TK = \{(t, k) \mid$ slot k is in period $t\}$. We define T_k as the time at which slot k ends and SL_k as the length of slot k. We fix some T_k to be the period ends with the upper bound of T_K being H. Denoting T_0 as the end of slot $k = 0$.

$$T_k = T_{(k-1)} + SL_k \qquad\qquad T_0 = 0,\ 0 < k \leq K \qquad\qquad (1)$$

In the following, sets $BJ = \{(b, j) \mid$ blender b can feed product tank $j\}$, $BP = \{(b, p) \mid$ blender b can process product $p\}$, and $JO = \{(j, o) \mid$ product tank j may deliver order $o\}$.

3.1. Blending and Storage

We use a dummy product tank ($j = 0$) to model the idle status of a blender. Thus, we have J real product tanks (j = 1, 2, …, J) and one ($j = 0$) dummy product tank. We define binary variables v_{bjk}, and u_{jpk}, and 0-1 continuous variables ue_{jk}, x_{bpk}, and xe_{bk}:

$$v_{bjk} = \begin{cases} 1 & \text{If blender } b \text{ feeds product tank } j \text{ in slot } k \\ 0 & \text{Otherwise} \end{cases} \qquad u_{jpk} = \begin{cases} 1 & \text{If product tank } j \text{ holds product } p \text{ in slot } k \\ 0 & \text{Otherwise} \end{cases}$$

$$ue_{jk} = \begin{cases} 1 & \text{If tank } j \text{ switches products at the end of } k \\ 0 & \text{Otherwise} \end{cases} \qquad x_{bpk} = \begin{cases} 1 & \text{If blender } b \text{ processes product } p \text{ during } k \\ 0 & \text{Otherwise} \end{cases}$$

$$xe_{bk} = \begin{cases} 1 & \text{If blender } b \text{ ends the current run in slot } k \\ 0 & \text{Otherwise} \end{cases}$$

Each blender must feed exactly one product tank (real or dummy) in each slot.

$$\sum_{j=0}^{J} v_{bjk} = 1 \qquad\qquad (b, j) \in BJ,\ 0 < k \leq K \qquad\qquad (2)$$

3.2. Run Length

We define RL_{bk} as the length of the current run of blender b at the end of slot k, if the run does not end during slot k, and zero otherwise. Thus, $RL_{b0} = 0$, if a run has ended at time zero, otherwise it is the current run length at time zero. To compute RL_{bk}, we write,

$$RL_{bk} \leq RL_{b(k-1)} + SL_k \qquad\qquad 0 < k \leq K \qquad\qquad (3)$$

$$RL_{bk} \leq H(1 - xe_{bk}) \qquad\qquad 0 \leq k \leq K \qquad\qquad (4)$$

Then, to ensure a minimum length (RL_{bp}^L) for each blend run, we demand,

$$RL_{b(k-1)} + SL_k + \left\{ \max_p \left(RL_{bp}^L \right) \right\} (1 - xe_{bk}) \geq \sum_{p=1}^{P} RL_{bp}^L x_{bpk} \qquad (b, p) \in BP,\ 0 < k \leq K \qquad (5)$$

3.3. Variable Blending Rate

If blender b is not idle during slot k, then Q_{bk} must be limited by the maximum (F_b^U) and minimum (F_b^L) of blender b.

$$Q_{bk} \leq F_b^U \cdot SL_k \qquad\qquad 0 < k \leq K \qquad\qquad (6a)$$

$$Q_{bk} + F_b^L \cdot H \cdot (v_{b0k} + xe_{bk}) \geq F_b^L \cdot SL_k \qquad\qquad 0 < k \leq K \qquad\qquad (6b)$$

Eq. 6 allows the blending rate to vary from slot to slot during a run. Normally, this is not done in practice. However, enforcing this makes the formulation nonlinear and nonconvex. Therefore, we have decided to deal with this issue later.

3.4. Order Delivery

We define one binary variable (z_{jok}) as follows to denote order delivery operation.

$$z_{jok} = \begin{cases} 1 & \text{If product tank } j \text{ is delivering order } o \text{ during slot } k \\ 0 & \text{Otherwise} \end{cases}$$

If tank j is not delivering o during k, then the delivery amount (DQ_{jok}) must be zero:

$$DQ_{jok} \leq TQ_o \cdot z_{jok} \qquad\qquad (j, o) \in \boldsymbol{JO}, \; 0 < j \leq J, \; 0 < k \leq K \qquad\qquad (7)$$

Eq. 7 allowed order delivery to be intermittent from a tank. We propose a simple procedure later to correct this situation.

3.5. Inventory Balance on Component

Within any period, the feed rate (F_{it}) of i from upstream units is constant, so the inventory of component i at the end of slot k [V_{ik} ($V_i^L \leq V_{ik} \leq V_i^U$)] is:

$$V_{ik} = V_{i(k-1)} + \sum_{t \in \mathbf{TK}} F_{it} SL_k - \sum_{b=1}^{B} q_{ibk} \qquad\qquad 0 < k \leq K \qquad\qquad (8)$$

3.6. Objective Function

The objective is to minimize the total operating cost including material (component), transition and demurrage costs.

$$\text{Minimize } TC = \sum_{i=1}^{I} \sum_{b=1}^{B} \sum_{k=1}^{K} c_i \cdot q_{ibk} + \sum_{b=1}^{B} \sum_{k=1}^{K-1} CB_b \cdot xe_{bk} + \sum_{j=1}^{J} \sum_{k=1}^{K-1} CT_j \cdot ue_{jk} + \sum_{o=1}^{O} DM_o \cdot d_o \qquad (9)$$

where, c_i is the price (\$ per unit volume) of component i, CB_b is the cost (\$ per occurrence) of transition on blender b, CT_j is the cost (\$ per occurrence) of transition in product tank j, and DM_o is the demurrage cost (\$ per unit time) of order o.

This completes our multi-period model (MPM), which comprises eqs. 1-9 and other equations such as product qualities, and inventory balance on product tank, which are not presented here. As mentioned before, it allows the blending rate to vary from slot to slot and order delivery to be discontinuous, which is undesirable in practice. Therefore, we need a procedure to adjust the solution from MPM to obtain a realistic schedule.

4. Schedule Adjustment

The optimal solution from MPM gives us the values of x_{bpk}, xe_{bk}, v_{b0k}, SL_k, and Q_{bk}. We use [x_{bpk}], [xe_{bk}], [v_{b0k}], [SL_k], and [Q_{bk}] respectively to denote their optimal values. The run lengths (CRL_{bk}) and volumes (CCQ_{bk}) are computed by,

$$CRL_{bk} = CCQ_{bk} = 0 \qquad\qquad \text{if } [xe_{bk}] = 1 \qquad\qquad (10a,b)$$

$$CRL_{bk} = CRL_{b(k-1)} + [SL_k] \qquad\qquad \text{if } [xe_{bk}] = 0 \qquad\qquad (11a)$$

$$CCQ_{bk} = CCQ_{b(k-1)} + [Q_{bk}] \qquad\qquad \text{if } [xe_{bk}] = 0 \qquad\qquad (11b)$$

Then, we compute the total volume (TCQ_{bk}) processed by a blender in a run as:

$$TCQ_{bk} = 0 \qquad\qquad \text{if } [xe_{bk}] = 0 \qquad\qquad (12a)$$

$$TCQ_{bk} = CCQ_{b(k-1)} + [Q_{bk}] \qquad\qquad \text{if } [xe_{bk}] = 1 \qquad\qquad (12b)$$

The blending rate (R_{bk}) for each blending run at the slot where it ends is computed:

$$R_{bk} = \max\left(F_b^L, \frac{CCQ_{b(k-1)} + [Q_{bk}]}{CRL_{b(k-1)} + [SL_k]} \right) \text{ for } k \text{ with } [xe_{bk}] = 1 \,\&\, [v_{b0k}] = 0 \qquad (13)$$

Then, we set R_{bk} for all slots within each run to be the same as the one computed above. Now, to obtain a realistic schedule with the constant blend rates computed above, we fix x_{bpk}, xe_{bk}, and v_{b0k}. This allows us to fix, remove, or change some variables and constraints in MPM. For instance, eq. 6 becomes:

$$Q_{bk} = 0 \qquad\qquad \text{for } (b, k) \text{ with } [v_{b0k}] = 1 \qquad\qquad (14a)$$

$$Q_{bk} = R_{bk} SL_k \qquad\qquad \text{for } (b, k) \text{ with } [xe_{bk}] = [v_{b0k}] = 0 \qquad\qquad (14b)$$

$$Q_{bk} \le R_{bk} \cdot SL_k \qquad\qquad \text{for } (b, k) \text{ with } [xe_{bk}] = 1 \,\&\, [v_{b0k}] = 0 \qquad\qquad (14c)$$

The revised model (RMPM) comprises eqs. 1-2, 7-9, 14 and other equations, whose solution ensures that blending campaigns have constant blend rates that are within the limits on the blending rates and minimum run lengths at the same time. The schedule from RMPM may still show intermittent delivery of orders. When the delivery is over contiguous slots, then this can be easily revised by simply delivering at a constant rate until the entire order, which is distributed over contiguous slots, is fully delivered.

The complete procedure is described in brief as follows. We first solve MPM model which allows the blending rates to vary from slot to slot. Then, we compute CRL_{bk}, CCQ_{bk}, and blending rate R_{bk} for each run. Fixing x_{bpk}, xe_{bk} and v_{b0k}, we solve RMPM model to obtain a realistic schedule which ensures constant blending rate in a run. Finally, we correct intermittent delivery of orders to obtain final schedule.

Table 1 Solution statistics of various algorithms/codes for MPM for Examples 1-14

Ex	Order	Algorithm	Discrete Variables	Continuous Variables	Constraints	Total CPU Time (s)	Cost (k$)
1	5	DICOPT	130	334	1393	78.7	5149.73
		BARON	130	334	1393	14400[*]	5149.73
		Ours	130	328	1384	12.6	5149.73
2	10	DICOPT	329	735	3609	798	3658.11
		BARON	329	735	3609	14400[*]	3678.11
		Ours	329	729	3600	137	3658.11
4	15	DICOPT	486	883	4849	4074	4576.67
		BARON	486	883	4849	14400[*]	4717.13
		Ours	486	877	4840	800	4556.67
8	20	DICOPT	1159	1797	12878	14400[*]	12495.55
		BARON	1159	1797	12878	14400[*]	9492.56
		Ours	1159	1779	12848	10821	8329.13
10	25	DICOPT	1685	2458	19006	14400[*]	N/A
		BARON	1685	2458	19006	14400[*]	18223.18
		Ours	1685	2436	18968	13215	11649.17
14	45	DICOPT	4830	6493	57756	118800[*]	N/A
		BARON	4830	6493	57756	118800[*]	N/A
		Ours	4830	6439	57657	118800[*]	19207.48

Note: CPU time limit for MIP of DICOPT is set at 10800 s for Exs 1-10, 36000 s for Ex 11, and 108000 s for Exs 12-14
CPU time limit for MPM of ours is set at 10800 s for Exs 1-10, 36000 s for Ex 11, and 108000 s for Exs 12-14
CPU time limit for RMPM of ours is set at 3600 s for Exs 1-10, and 10800 s for Exs 11-14
Total CPU time limit of DICOPT, BARON, and ours is set at 14400 s for Exs 1-10, 46800 s for Ex 11, and 118800 s for Exs 12-14
* Reached total CPU time limit   N/A: No feasible solution

5. Detailed Evaluation

Recall that forcing the blending rate to be constant during a run makes the formulation nonlinear and nonconvex, which can be solved using commercial MINLP solvers such as BARON and DICOPT. Our adjustment procedure obviated the need to solve

MINLPs. We solve 14 examples with widely varying sizes, structures, scale, and complexity to evaluate our formulation and solution approach. Table 1 shows the solution statistics only for Examples 1-2, 4, 8, 10, and 14. For Example 1, our procedure needs only 12.6 CPU s for the optimal solution of 5149.73 k$, but DICOPT needs 78.7 CPU s and 14400 CPU s for BARON. For Example 2, our procedure obtains the optimal solution of 3658.11 k$ within 137 CPU s, but DICOPT needs 798 CPU s. BARON gets a worse solution of 3678.11 k$ after 14400 CPU s. For Example 4, our approach obtains the optimal solution of 4556.67 k$, but DICOPT and BARON do not. For Examples 3 and 5 (not in Table 1), DICOPT does get the optimal solutions, but requires an order of magnitude longer solution times compared to our procedure. For instance, DICOPT takes 531 CPU s for Example 3 versus only 83 CPU s for our procedure. Interestingly, BARON also reaches the optimal solution for Example 3, but needs 14400 CPU s. For the remaining examples (Examples 6-14), our approach always obtains better solutions than both DICOPT and BARON within the allocated CPU time. For instance, our approach finds a solution of 8329.13 k$ for Example 8, while DICOPT obtains 12495.55 k$, and BARON gets 9492.56 k$. Moreover, our approach obtains a solution of 11649.17 k$ for Example 10, while BARON gets 18223.18 k$, and DICOPT cannot obtain a feasible solution.

6. Conclusion

We developed a slot-based multi-period continuous-time model for integrated scheduling of gasoline blending and distribution operations in a refinery and incorporated many real-life operating features and policies. We proposed an ingenious schedule adjustment procedure that requires only MILP solutions to ensure constant blending rate in a run. On 14 test problems of varying sizes and features, our proposed procedure was superior to commercial solvers (DICOPT and BARON). Much further work is needed, as our model still cannot optimally solve truly large problems involving more than 30 orders within reasonable time.

References

K. Glismann, G. Gruhn, 2001, Short-term scheduling and recipe optimization of blending processes, *Comput Chem Eng*., 25, 627-634.

Z. Y. Jia, M. Ierapetritou, 2003, Mixed-integer linear programming model for gasoline blending and distribution scheduling, *Ind. Eng. Chem. Res.*, 42, 825-835.

I. A. Karimi, C. M. McDonald, 1997, Planning and scheduling of parallel semicontinuous processes. 2. short-term scheduling, *Ind Eng Chem Res*., 36, 2701-2714.

J. Li, W.K. Li, I. A. Karimi, R. Srinivasan, 2007, Improving the robustness and efficiency of crude scheduling algorithms, *AIChE J.*, 52, 10, 2659-2680.

Y. Liu, I. A. Karimi, 2007, Scheduling multistage, multiproduct batch plants with nonidentical parallel units and unlimited intermediate storage, *Chem Eng Sci.*, 62, 1549-1566.

C. A. Mendez, I. E. Grossmann, I. Harjunkoski, P. Kabore, 2006, A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations, *Comput Chem Eng.*, 30, 614-634.

J. M. Pinto, M. Joly, L. F. L. Moro, 2000, Planning and scheduling models for refinery operations, *Comput Chem Eng.*, 24, 2259-2276.

P. C. P. Reddy, I. A. Karim, R. Srinivasan, 2004a, A new continuous-time formulation for scheduling crude oil operations, *Chem. Eng. Sci.*, 59, 1325-1341.

P. C. P. Reddy, I. A. Karim, R. Srinivasan, 2004b, A novel solution approach for optimizing scheduling crude oil operations, *AIChE J.*, 50, 6, 1177-1197.

NARX-Model-Based Control (NARX-MBC) for Citronellyl Laurate Esterification Reactor

S.A. Zulkeflee, N.Aziz*

School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang Malaysia
*Phone: +604 -5996457, Fax: +604 -5941013, Email: chnaziz@eng.usm.my

Abstract

Model-based control nowadays appears to be a very promising control strategy for various processes. However, this type of control strategy requires an accurate, low complexity, easy identifiability, structural flexibility and if possible, invertibility type of models. In this paper, a non-linear autoregressive exogenous-model-based control (NARX-MBC) has been designed and implemented to a batch citronellyl laurate esterification reactor. Multi-input-single-output (MISO) model has been developed to representing the process using NARX modelling approach. The performance of the developed NARX-MBC is evaluated and compared with a conventional PID controller. Overall, it is observed that the former has outperformed the latter.

Keywords: NARX, model based control, model identification, batch esterification.

1. Introduction

Batch reactors are frequently used in chemical, petrochemical or biochemical industries, for various quality products. Bioreactors are complex, as they generally involve subsystems that have numerous internal states, and require a number of parameters with highly nonlinear inter-relationships. As a result, the modeling and control of bioreactors is difficult and due to this reason, the application of model-based control strategies often gives better control performance (Arpornwichanop et. al., 2005).

In recent years, several researchers have implemented model based control strategies in a batch reactor (Konakom et. al., 2008; Zhang, 2008; Oliveira, 2004; Preub et. al., 2003; Aziz et. al., 2000). However, to the best of author's knowledge, there is no work carried out to implement model based control strategies for batch esterification process. In this work, NARX-Model-Based-Control (NARX-MBC) is designed and applied to control the reactor temperature of batch citronellyl laurate esterification process. Ability to track the desired temperature will lead the process reaching to the required ester conversion as demanded by the customers.

Nonlinear autoregressive with exogenous inputs (NARX) model is able to provide a powerful representation for time series analysis, modeling and prediction due to its strength to accommodate the dynamic, complex and nonlinear nature of real time series applications (Harris and Yu, 2007; Mu et. al.,2005). Due to such advantages, in this work NARX model has been chosen to represent the process under consideration. NARX model has been used as identification model that can capture the non-linear dynamics relating the inputs to the output of the system.

For control purposes, the developed NARX model is embedded in the internal model control strategy and known here as NARX-Model-Based-Control (NARX-

MBC). Finally, NARX-MBC performance is evaluated and compared with a conventional PID controller.

2. Citronelly laurate esterification process

In this work, the citronellyl laurate esterification process reaction (Aziah et. al., 2006) taking place in the jacketed batch reactor has been considered (Figure 1). The specification of reactor chosen in this study is similar to the batch reactor rig available at the Process Control Laboratory, School of Chemical Engineering, Universiti Sains Malaysia. This process is a well-mixed, heterogeneous with exothermic reaction scheme. Figure 2 shows the schematic that represent esterification of citronellyl laurate where citronellol and lauric acid are the reactants. Citronellyl laurate (ester) is the desired product while water is an unwanted byproduct. The immobilized *candida rugosa* lipase on to amberlite MB-1 is used as catalyst for the synthesis of the citronellyl laurate.
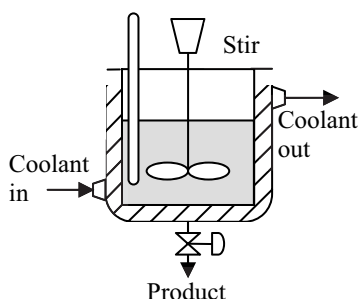
Figure 1: Schematic diagram of jacketed batch reactor
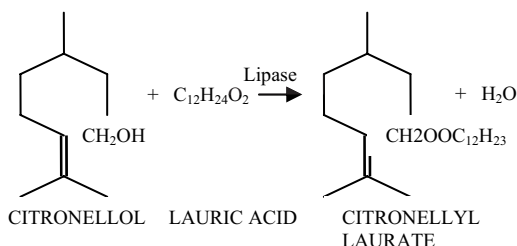
Figure 2: Schematic represents esterification of citronellyl laurate

The kinetic data for the model is obtained from analysis of the initial rate data and progress curve data and the reaction conducted with immobilized lipase follows an ordered bi-bi mechanism with dead-end of lauric acid (Garcia et. al., 2000).

3. NARX model

NARX model is a nonlinear generalization of a well known ARX model, which constitute a standard tool in linear black-box identification. A NARX model $\mu(\theta)$ is characterized by the one-step ahead predictor:

$y(t|\theta)=F(x(t))$

where the vector $x(t)=[y(t-1),...Y(t-n),u(t-1),...u(t-m)]$ represents the memory of the model. Since the function F is, to a large extent, unknown it is approximated by an expression in a given function basis. The function $F(.)$ can include both linear and nonlinear functions of $x(t)$ as shown in Figure 3. Data for the input $u(t)$ and output $y(t)$ in this process is generated from the validated first principle models (Zulkeflee and Aziz, 2008) with operating conditions as shown in Table 1.

Figure 3: NARX blocks diagram.

Table 1: Operating Conditions and Calculated Parameters

Parameter	Values	Parameter	Values	Parameter	Values
A_{Ac}	18.20 L mol/s	Cp_{Ac}	420.53J/mol K	α	1
A_{Al}	24.04 L mol/s	Cp_{Al}	235.27J/mol K	β	1
A_i	0.32 L mol/s	$Cp_{Es\backslash}$	617.79J/mol K	U	2.86J/sm^{2}K
E_{Ac}	-105.40J mol/K	Cp_w	75.40J/mol K	A	0.077 m^2
E_{Al}	-66.09 J mol/K	V	1.5 L	ΔH_{rxn}	19.84 kJ
E_i	-249.94J mol/K	V_j	0.8 L	T_{ji}	294 K

NARX model use the wavelet network as a nonlinearity estimator. The wavelet network function is based on the following function expansion:

$$F(x)=(x-r)\,PL + a_{s1}f((b_{s1}(x-r))Q - c_{s1}) + a_{sn}f((b_{sn}(x-r))Q - c_{sn})+\ldots$$
$$a_{w1}g((b_{w1}(x-r))\,Q - c_{w1}) + a_{wn}wg((b_{wn}w(x-r))Q - c_{wn}) + d$$

From the above equation, f is a scaling function and g is the wavelet function. P and Q are m-by-p and m-by-q projection matrices, respectively. The projection matrices P and Q are determined by principal component analysis of estimation data. Usually, $p=m$. If the components of x in the estimation data are linearly dependent, then $p<m$. The number of columns of Q, q, corresponds to the number of components of x used in the scaling and wavelet function. Its structure and parameters are tabulated in Table 2.

Table 2: NARX model obtained with model analysis

Scaling Translation, cs_k	Scaling Coefficient, as_k	Scaling Dilation, bs_k
$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & -1 \\ 0 & -1 & -2 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 1 & 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & -1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 0 & -1 & 0 & -2 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 & -1 \\ 0 & 1 & -1 & -1 & 1 & -1 & -1 & -1 & 2 \\ 0 & 1 & -1 & -1 & 0 & 1 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & -2 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & -2 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 2 & 2 \\ 1 & -1 & 1 & 0 & 0 & -2 & 2 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 1 & 1 & 0 & 1 & -1 & 1 & 0 \\ 1 & -2 & 1 & 1 & 0 & -2 & 1 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 1 & -2 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -0.0260 \\ 0.0382 \\ -0.0579 \\ 0.0671 \\ 0.0336 \\ 0.0875 \\ 0.0468 \\ 0.0306 \\ 0.0670 \\ -0.0220 \\ -0.0298 \\ 0.0259 \\ 0.0034 \\ 0.0268 \\ 0.0009 \\ -0.0290 \\ 0.0303 \\ -0.0170 \\ 0.0011 \\ -0.0014 \end{pmatrix}$	[1;1;1;1;1;1;1;1;1;1;1;1;1;1;1; 1;1;1;1;1] Output Offset, d 320.9573 Regressor Mean, r [320.9097 320.85797 320.8021 0.0085 0.0084 0.0084 293.1141 293.1074 293.10076] Wavelet Dilation, bw_k [1;1;1;1;1;1;1;1;1;1;1;1;1;1;1; 1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1 ;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1; 1;1;1;1;1;1;1;1;1;1; 1;1;1;1;1;1;1;1;1]

R-value for the state-space, ARX and NARX model validation achieved are 0.4849, 0.8382 and 0.9367 respectively. It is show that NARX performed better than the other

two models and thus will be used in the control section. Figure 4 shows the output profile for validation between identified NARX model and validation data.

Figure 4: Simulation of model output compared against a validation data set

4. NARX Model-Based Control (NARX-MBC)

The IMC structure (Garcia & Morari, 1982) given in Figure 5 is a central to our discussions on designing the controller. Its conceptual usefulness lies in fact that it allows us to concentrate on the controller design without having to concern with control system stability provided that the process model $p^{\sim}(s)$ is a NARX model representation of a stable process. In this work, the process $p(s)$ is represented by the first principle model of the esterification batch reactor. The controlled variable of the system considered is temperature of the reactor (Tr) whereas the manipulated variable is cooling water flow (Fj). Meanwhile, the cooling water temperature (Tj) is considered as disturbance. The controller $q(s,\varepsilon)$ of this system is the inverse model of the process $p^{-1}(s)$.
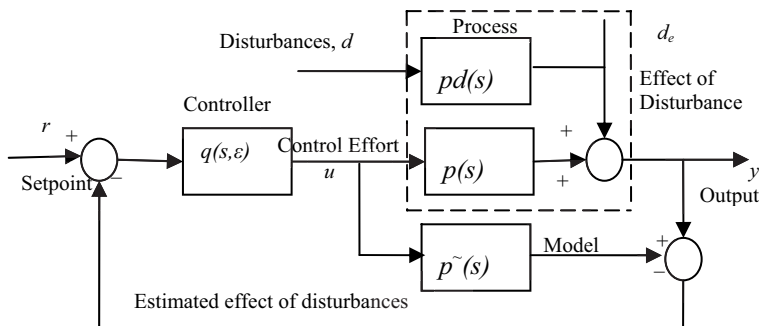
Figure 5: The IMC system

The inverse model is derived from first order plus time delay that developed using NARX model. Model parameters at different steady state points in esterification reaction are identified by Cohen Coon method (Seborg et . al. 2004) and the results are tabulated in Table 3. Worst case of model parameters of the process is chosen to develop first order model (Bhapa et. al. 2005). In this method the step test with largest process gain (i.e. $K_p = 82405$) and smallest time constant (i.e. $\tau = 5.2047$) were chosen as

the worst case for the process. The primary controller takes the form $q(s,\varepsilon)=p\text{-}1(s)f$, where f is a user specified low-pass filter;

$$F(s,\tau_c)=1/(\tau_c+1)^{\,r}$$

where r is sufficiently positive integer in order to guarantee that the IMC controller is proper. The r =1 and τ_c=0.01.

$$q(s,\varepsilon) = \frac{5.2047\ s + 1}{82405\ (\dfrac{1}{0.01s + 1})}$$

Table 3: Identified Model Parameters at different steady state points

Cases (L/min)	0-0.002	0.002-0.004	0.012-0.014	0.014-0.016	0.016-0.018
K_p	82405	81700	22490	19610	17633
Tp	26.863	20.193	5.7052	5.2047	7.8423

5. Results and Discussion
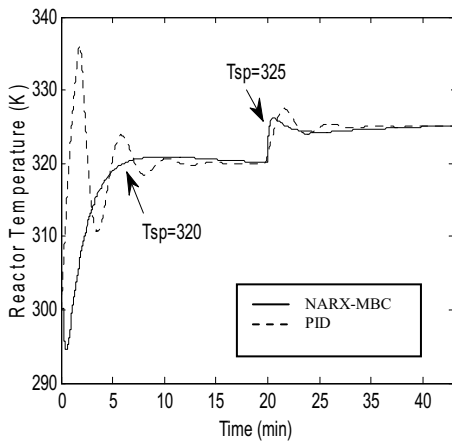
Figure 6: Set-point change for NARX-MBC and PID controller
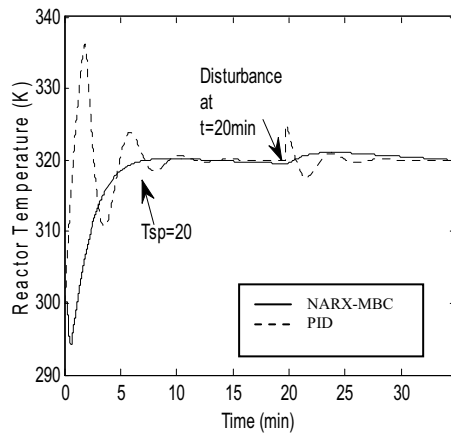
Figure 7: Disturbance rejection for NARX-MBC and PID Controllers

The responses for set point tracking and disturbance rejection for both NARX-MBC and PID controllers are shown in Figures 6 and 7 respectively. For both cases, NARX-MBC found to give faster and smoother closed loop compared to PID. The result also show that PID controller has difficulty in following the desired temperature profile and lead to a poor control performance. As shown in Figures 6 and 7, PID controller found to produce sluggish damping oscillations response for both set point change and disturbance rejection. On the other hand, NARX-MBC produce response with low amount of overshoot and less oscillate. NARX-MBC also able to settle faster compared to PID. Overall, NARX-MBC provides superior performance in set-point tracking and disturbance rejection than PID.

6. Conclusions

NARX-MBC has been designed and applied to the batch esterification process to produce citronellyl laurate. MISO NARX model to represent the process has been developed and the validation result showed that it was more than 93% accurate. The

validated NARX was then embedded in the control strategy proposed. The results showed that the NARX-MBC can cope with both set point change and disturbance rejection. It was then compared with PID controller and the former found to outperform the latter.

Acknowledgements

References

A. Arpornwichanop, P. Kittisupakorn and I.M Mujtaba, 2005, On-line dynamic optimization and control strategy for improving the performance of batch reactors, Chemical Engineering and Processing 44, 101-114.

S.N. Aziah, A.H. Kamaruddin and W.S Long, 2006, Studies of reaction parameters on synthesis of citronellyl laurate ester via immobilized *Candida rugosa* lipase in organic media, Bioprocess and Biosystems Engineering 29, 253-260.

N. Aziz, M.A. Hussain and I.M. Mujtaba, 2000, Performance of different types of controllers in tracking optimal temperature profiles in batch reactors, Computers and Chemical Engineering 24, 1069-1075

P.K. Bhaba, S. Sathishbabu, A. Asokan and T. Karunanithi, 2005, Real Time Implementation of Wiener Model PI (WMPI) Controller in a Conical Tank Liquid Level Process,ICCBPE/SOMChe, 297-302.

T.J. Harris and W. Yu, 2007, Controller assessment for a class of nonlinear systems, Journal of Process Control 17 607-619

C.E. Garcia and M. Morari, 1982, Internal model control: A unifying review and some new results. Ind. Eng. Chem. Process Des. Dev. 21 (2), 308-323

T. Garcia, A. Coteron, M. Martinez and J. Aracil, 2000, Kinetic model for the esterification of oleic acid and cetyl alcohol using immobilized lipase as catalyst, *Chemical Engineering Science, 55*, 1411-1423

K. Konakom, P. Kittisupakorn and I.M. Mujtaba, 2008, Batch control improvement by model predictive control based on multiple reduced-models, Chemical Engineering Journal.

J. Mu, D. Rees and G.P. Liu, 2005, Advanced controller design for aircraft gas turbine engines. Control Engineering Practice 13 1001–1015

R. Oliveira, 2004, Combining first principles modelling and artificial neural networks: a general framework. Computers and Chemical Engineering 28, 755-766.

K. Preub, M.V. Lann, M. Cabassud and G. Anne-Archard, 2003, Implementation procedure of an advanced supervisory and control strategy in the pharmaceutical industry, Control Engineering Practice, 11, 1449-1458.

D.E. Seborg, T.F. Edgar and D.A. Mellichamp, 2004, Process Dynamics and Control, 2nd edition.

J. Zhang, 2008, Batch-to-batch optimal control of a batch polymerisation process based on stacked neural network models. Chemical Engineering Science 63 1273– 1281

S.A. Zulkeflee and N. Aziz, 2008, Modeling and simulation of citronellyl laurate esterification in batch reactor. In Proceedings of Regional Symposium on Chemical Engineering (RSCE) in conjunction with Symposium of Malaysian Chemical Engineers, Kuala Lumpur, Malaysia (SOMChE), Vol II, 955-960.

1575

Nonlinear Model Predictive Control of a Distillation Column Using NARX Model

K. Ramesh, S. R. Abd Shukor and N. Aziz*

School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia, Seri Ampangan, 14300 Nibong Tebal, Pulau Pinang, Malaysia
* Phone: +604 -5996457, Fax: +604 -5941013, Email: chnaziz@eng.usm.my

Abstract

Distillation column is an important process unit in petroleum refining and chemical industries, and needs to be controlled close to optimum operating conditions because of economic incentives. Nonlinear model based control (NMPC) scheme is one of the best options to be explored for proper control of distillation columns. In this work, NMPC scheme using sigmoidnet based nonlinear autoregressive with exogenous inputs (NARX) model has been developed to control distillation column The Unscented Kalman Filter (UKF) was used to estimate the state variables in NMPC and the nonlinear programming (NLP) problem was solved using sequential quadratic programming (SQP) method. The closed loop control studies have indicated that the NARX NMPC performed well in disturbance rejection and set point tracking.

Keywords: Distillation column, NARX model, Nonlinear model predictive control.

1. Introduction

Model Predictive Control (MPC) is an important advanced control technique which can be used for difficult multivariable control problems (Goodwin *et al.*,2001). The current generation of commercially available MPC technology is based on linear dynamic models, and is refereed by the generic term linear model predictive control (LMPC). Many processes such as high purity distillation column, multi-grade polymer reactors are sufficiently nonlinear to preclude the successful application of LMPC technology. This has led to the development of nonlinear model based controllers such as nonlinear model predictive control (NMPC) in which more accurate nonlinear model is used for process prediction and optimization.

Many authors have studied the performance of NMPC to control distillation using different nonlinear models namely semi-rigorous reduced order model (Maiti and Saraf, 1995), NARX model (Sriniwas *et al.*, 1995), Hammerstein model (Fruzzetti *et al.*, 1997), Recurrent Dynamic Neuron Network (RDNN) model (Shaw and Doyle III, 1997) and grouped neural networks (GNN) model (Ou and Rhinehart, 2003). Foss *et al*. (1998), in their case study on process modeling in Germany and Norway concluded that despite the commercially available modeling tools, the effort spent for all kinds of modeling activities is the most time consuming step in an industrial project where model based process engineering techniques are applied.

The NMPC problem formulation involves online computation of a sequence of manipulated inputs which optimize an objective function and satisfy process constraints. NMPC requires online solution of a nonlinear program (NLP) at each iteration. The solution of such NLP problems can be very time consuming, especially for large scale systems. An additional complication is that the optimization problem generally is

nonconvex because the nonlinear model equations are posed as constraints (Cannon,2004). Consequently, NLP solvers designed for convex problems may converge to local minima or even diverge. So it is necessary to find out an improved solution algorithm for nonconvex NLP problems.

In this paper, two multiple-input-single-output (MISO) sigmoidnet based NARX models were used to model the dynamics of the distillation column. An equilibrium model for distillation column was used as plant model in nonlinear system identification and in NMPC.

2. Sigmoidnet based NARX model

Two multiple-input-single-output (MISO) NARX models are developed in this study to model the dynamics of the distillation column. The reason for using two MISO models instead multiple-input-multiple-output (MIMO) model is that the MISO models provide better prediction compared to MIMO model (Eskinat *et al.*,1991) The first MISO model using reflux flow rate (L) and reboiler heat load (Q_R) as inputs, and top product composition (x_D) as output, while, the second MISO model using reflux flow rate (L) and reboiler heat load (Q_R) as inputs, and bottom product composition (x_B) as output. The models consisted of parallel combination of nonlinear and linear blocks.

In both the MISO NARX models used this work, sigmoidnet function with 2 units is used as nonlinear regression function containing two past output regressors and two past input regressors (one from each input). The parameters of sigmoidnet based MISO NARX models were estimated using the *ident* function system identification toolbox version 7.0 in MATLAB. The iterative prediction-error minimization method discussed in Ljung (1999) was used to calculate the model parameters. The data generated from equilibrium model for distillation column discussed in Ramesh *et al*. (2005) was used for parameter estimation and model validation. The model structure, parameter estimation, validation and model analysis are discussed in detail by Ramesh *et* a.l (2008).

3. NMPC

The developed nonlinear wavenet based Hammerstein model is of the following form.

$$X(k+1) = F[X(k),U(k)] \tag{1}$$

$$Y(k) = h[X(k)] \tag{2}$$

where $X(k) = [x_1(k)\ x_2(k)]^T$; $U(k) = [u_1(k)\ u_2(k)]^T$; $Y(k) = [y_1(k)\ y_2(k)]^T$

x_1 and x_2 are n-dimensional vector of state variables, u_1 (reflux flow rate) and u_2 (reboiler heat load) are m-dimensional vectors of manipulated input variables, and y_1 (top product composition) and y_2 (bottom product composition) are p-dimensional vector of controlled output variables. In this work, two separate MISO models were developed (one for each output) instead of using a MIMO model. The optimization problem is given by

$$\min_{U(k\backslash k),U(k+1\backslash k)...,U(k+M-1\backslash k)} J = \phi[\boldsymbol{Y}(k+P\backslash k)] + \sum_{j=0}^{P-1} L[Y(k+j\backslash k), U(k+j\backslash k), \Delta U(k+j\backslash k)] \tag{3}$$

where $U(k+j\backslash k)$ is the input $U(k+j)$ calculated from information available at time k, $Y(k+j\backslash k)$ is the output $Y(k+j)$ calculated from information available at time k, $\Delta U(k+j\backslash k) = U(k+j-1\backslash k) - U(k+j\backslash k)$, M is the control horizon, P is the prediction horizon and φ and L are nonlinear functions of their arguments. The functions φ and L can be chosen to satisfy wide variety of objectives and in this study, the quadratic functions of the following form in considered:

$$L = [Y(k+j \setminus k - Y_s(k)]^T Q[Y(k+j \setminus k - Y_s(k)]$$
$$+ [U(k+j \setminus k - U_s(k)]^T R[U(k+j \setminus k - U_s(k)] \qquad (4)$$
$$+ \Delta U^T(k+j \setminus k)S\Delta U(k+j \setminus k)$$

$$\phi = [Y(k+P \setminus k) - Y_s(k)]^T Q[Y(k+P \setminus k) - Y_s(k)] \qquad (5)$$

where $U_s(k)$ and $Y_s(k)$ are steady-state targets for U and Y respectively, and Q, R and S are positive-definite weighing matrices. The principal controller tuning parameters are M, P, Q, R, S and the sampling period Δt.

The reflux flow rate bounds are set to be [0.1, 0.75] l/min. The lower bound for reflux flow rate 0.1 l/min was meant to keep the input physically meaningful, namely, the reflux flow rate should be positive and have some minimum value. The upper bound of reflux flow rate is approximately 150% of the nominal capacity, which would seldom occur in operation. The reboiler heat load bounds are set to be [0, 15] kW. The lower bound meant that the reboiler heat load should not be negative, whereas upper bound 15 kW was the maximum heater capacity of the reboiler.

The top product composition bounds are set to be [0.5, 1]. The lower bound for output 0.5 was meant that the top product purity should not be less than 50%. The upper bound 1 was meant that the maximum value of top product purity is 100% and beyond that is practically not meaningful. The bottom product composition bounds are set to be [0, 0.5]. The lower bound for bottom product composition 0 was meant that the maximum value of bottom product purity is 100%. The upper bound of bottom product composition 0.5 was meant that the bottom product purity should not be less than 50%. The desired product purity is achieved by solving the nonlinear optimization problem subject to the following inequality constraints.

$$U_{min} \le U(k+j \setminus k) \le U_{max}, \quad 0 \le j \le M-1 \qquad (6a)$$

$$\Delta U_{min} \le \Delta U(k+j \setminus k) \le \Delta U_{max}, \quad 0 \le j \le M-1 \qquad (6b)$$

$$Y_{min} \le Y(k+j \setminus k) \le Y_{max}, \quad 1 \le j \le P \qquad (6c)$$

In addition, the nonlinear model equations are posed as a set of following equality constraints:

$$X(k+j+1 \setminus k) = F[X(k+j \setminus k), U(k+j \setminus k)], \quad 0 \le j \le P-1 \qquad (7a)$$

$$Y(k+j \setminus k) = h[X(k+j \setminus k)], \quad 1 \le j \le P \qquad (7b)$$

where $X(k \setminus k) = X(k)$ if the state variables are measured. It is important to note that input constraints are hard constraints in the sense that they must be satisfied. Conversely, output constraints can be viewed as soft constraints because their violation is necessary to obtain a feasible optimization problem.

NMPC calculation requires measurements or estimates of the state variables and in the present work, UKF was used to estimate the state variables in the NMPC problem. In UKF, the state distribution is represented by a Gaussian Random Variables (GRV), which is specified using a minimal set of carefully chosen sample points. These sample points completely capture the true mean and covariance of the GRV, and when propagated through the true non-linear system, captures the posterior mean and covariance accurately to the third order of Taylor series expansion for any nonlinearity (Wan and Merwe,2000). Finally, the NMPC problem was solved using *fmincon* function in MATLAB optimization toolbox version 3.1.1which uses a sequential quadratic programming (SQP) method.

4. Closed loop control studies

A sampling interval of 1 min was chosen for closed loop control studies. The NMPC parameters M, P, Q, R, S and the sampling period Δt are chosen by repeated tuning and the final values are; sampling period $\Delta t = 1$ min, prediction horizon $P = 30$ time steps (30 min), control horizon $M = 6$ time steps (6 min), $Q = (1, 0.5)$, $R = (1,1)$ and $S = (1,1)$. The equilibrium model for distillation column discussed in Ramesh et $al.$ (2005) was used as the plant model
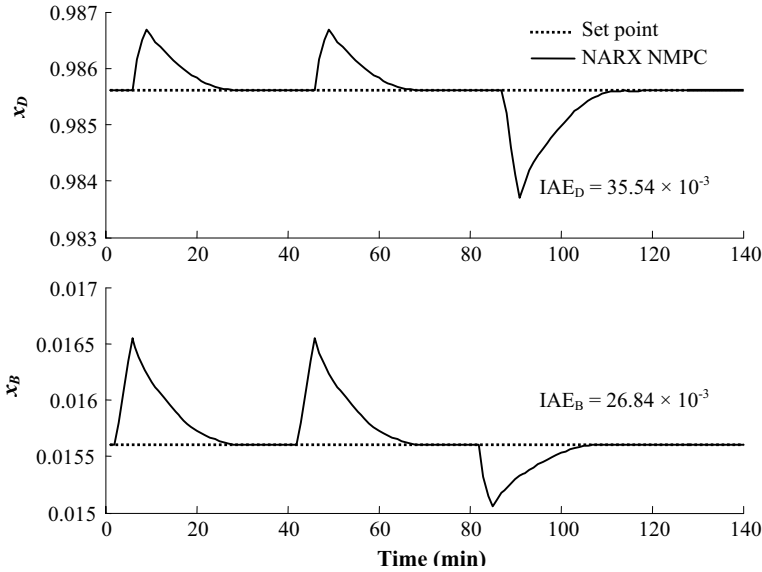
Figure 2: NARX NMPC responses of product compositions to a feed flow rate disturbance
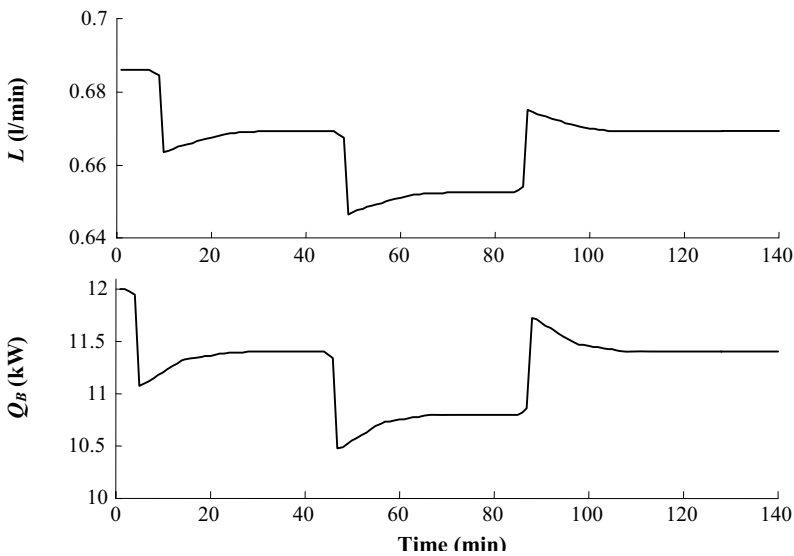
Figure 3: NARX NMPC responses of manipulated variables to a feed flow rate disturbance

The performance of the NMPC was studied by making three feed flow rate disturbances: a +20% increase at t = 0 min, again a +20% increase at t = 40 min and a -20% decrease at t = 80 min. The responses of product compositions for these feed flow rate disturbances are shown in Figure 2 along with the corresponding numerical values of IAE. It can be observed from the figure that NARX NMPC was reasonably rejected the feed flow rate disturbances. The corresponding responses of manipulated variables to feed flow rate disturbances are shown in Figure 3.
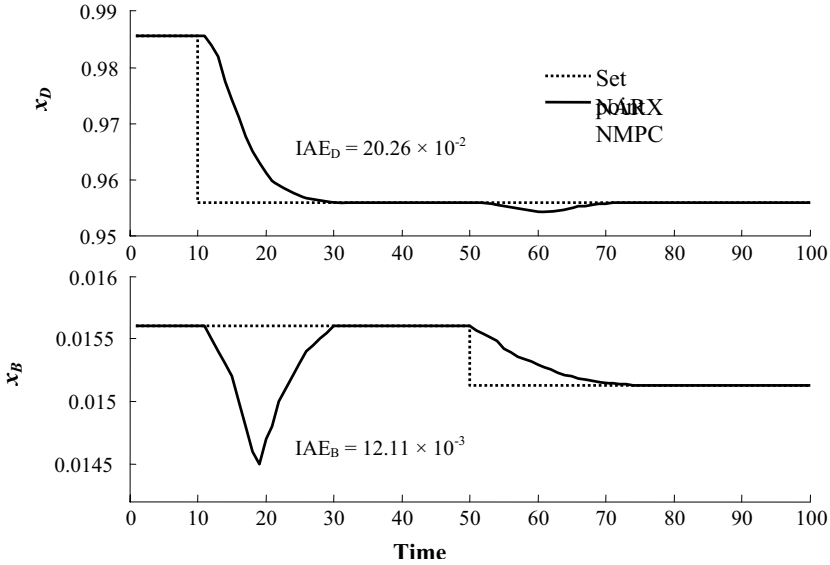
Figure 4: NARX NMPC responses of product compositions to -3% change in x_D at t =10 min and -3% change in x_B at t = 50 min
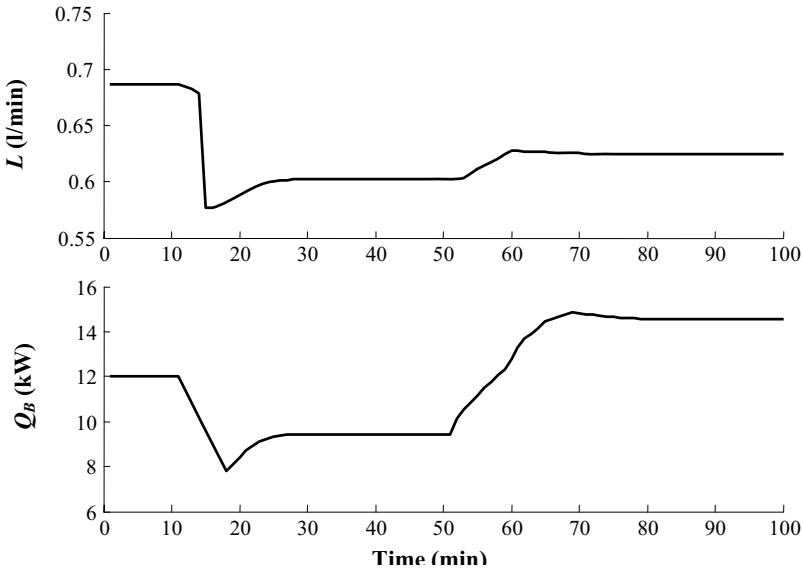
Figure 5: NARX NMPC responses of manipulated variables to -3% change in x_D at t =10 min and -3% change in x_B at t = 50 min

The performance of the NARX NMPC in set point tracking was studied by making -3% change in x_D at t = 10 min followed by -3% change in x_B at t = 50 min. The responses of product compositions for these set point changes are shown in Figure 4 and it can be noted from the figure that the NARX NMPC tracking the set point well. The corresponding responses of manipulated variables for set point changes are shown in Figure 5. It was noted that the reboiler should be operated very close to the upper limit of the Q_B in order to track the set point changes.

5. Conclusions

The sigmoidnet based NARX model NMPC was developed to control distillation column. The Unscented Kalman Filter (UKF) was used to estimate the state variables in NMPC and the NLP problem was solved using sequential quadratic programming (SQP) method. The closed loop control studies indicated that the developed NMPC technique performed well in controlling the distillation column by rejecting the disturbances in regulatory control and tracking the set point quickly in servo control.

Acknowledgement

References

M. Cannon 2004. Efficient nonlinear model predictive control algorithms. Annual Reviews in Control 28: 229-237.

E. Eskinat, S. H. Johnson and W. L. Luyben 1991. Use of Hammerstein models in identification of nonlinear systems. A.I.Ch.E. Journal 37(2): 255–268.

B. Foss, B. Lohmann and W. Marquadt 1998. A field study of the industrial modeling process. Journal of Process Control 8: 325–337.

K. P. Fruzzetti, A. Palazoglu and K. A. McDonald 1997. Nonlinear model predictive control using Hammerstein models. Journal of Process Control 7(1): 31-41.

G. C. Goodwin, S. F. Graebe and M. E. Salgado 2001. Control System Design. Upper Saddle River, NJ, Prentice Hall.

L. Ljung 1999. System Identification: Theory for the User. Upper Saddle River, NJ, Prentice Hall PTR.

S. N. Maiti and D. N. Saraf 1995. Adaptive dynamic matrix control of a distillation column with closed-loop online identification. Journal of Process Control 5(5): 315-327.

J. Ou and R. R. Rhinehart 2003. Grouped neural network model-predictive control. Control Engineering Practice 11(7): 723-732.

K. Ramesh, N. Aziz and S. R. Abd Shukor 2008. Development of sigmoidnet NARX model for a distillation column. Chemical Products and Process Modeling 3(2): Article 4.

K. Ramesh, M. Ramasamy, N. Aziz, S. R. Abd Shukor and M. Zailani Abu Bakar 2005. Comparison of rate- based and equilibrium models for multicomponent distillation. In Proceedings of International Conference on Chemical and Bioprocess Engineering 2005, Sabah, Malaysia.: 549-554.

A. M. Shaw and F. J. Doyle III 1997. Multivariable nonlinear control applications for a high purity distillation column using a recurrent dynamic neuron model. Journal of Process Control 7(4): 255-268.

G. R. Sriniwas, Y. Arkun, I. L. Chien and B. A. Ogunnaike 1995. Nonlinear identification and control of a high-purity distillation column: a case study. Journal of Process Control 5: 149-162.

E. A. Wan and R. V. D. Merwe 2000. The Unscented Kalman Filter for Nonlinear Estimation. In Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC), Lake Louise, Alberta, Canada., IEEE.

A Neighbour in Control Technique for Batch Process Monitoring

Carlos Alvarez, Adriana Brandolin, Mabel C. Sánchez

Planta Piloto de Ingeniería Química (UNS-CONICET), Camino La Carrindanga Km 7, Bahía Blanca 8000, Argentina

Abstract

In this work a Nearest in Control Neighbour (NICN) based strategy is presented for monitoring batch processes. The technique detects a fault if the T^2 statistic value exceeds the critical one for three successive observations. In this case, the identification of suspicious measurements is determined in terms of the distance between the current observation and its NICN. Furthermore a comparison between the NICN approach and the classical Multiway Principal Component Analysis (MPCA) is provided. Both techniques are applied to supervise the operation of a semibatch reactor for methyl-methacrylate polymerization. The analysis of results reveals that NICN strategy is suitable for fault detection of batch processes and has the advantage of avoiding ambiguous identifications when compared with classical techniques.

Keywords: Statistical Process Control, Multiway Principal Component Analysis, Original-variable Space.

1. Introduction

In the last few decades batch processes experienced a renaissance as products-on-demand and first-to-market strategies impel the need for flexible and specialized production methods. They are mainly devoted to the production of polymers, pharmaceuticals, foods, bio-chemicals and semiconductors.

Online process performance monitoring and product quality prediction in real time ensure safe and profitable operation of batch processes since they provide the opportunity to take corrective actions before the effects of excursions from normal operation ruin the batch.

Several multivariate statistical techniques have been proposed for online monitoring of batch processes. Multiway Principal Component Analysis (MPCA) and Multiway Partial Least Squares (MPLS) are some of the most commonly used in statistical process control. Using a model of the normal process behaviour in a reduced space of latent variables, these techniques monitor the process performance by comparing the time progression of the data with that of the average normal batch, which is obtained from the data base of normal operating conditions. A comprehensive discussion of PCA-based statistical techniques for monitoring batch processes can be found in the review work by Westerhuis et al. (2000).

Recently Alvarez et al. (2008) presented a novel approach for monitoring continuous processes in the original-variable space. The detection stage of the technique makes use of the T^2 statistic value to indicate the out of control state. In the identification stage, the distance between the Nearest In Control Neighbour (NICN) and the current observation is used to determine the contribution of each variable to the out of control state. Those variables whose distance measures exceed a certain threshold value are considered as

suspicious. The strategy has shown good performance for monitoring simple continuous processes.

In this work an extension of the NICN strategy to address batch monitoring is presented. Different threshold values are employed for identification purposes. The performance of the proposed technique is evaluated and compared with that provided by MPCA (Nomikos and MacGregor, 1994) for monitoring a polymerization batch reactor. The measures of performance are the number of false and missing alarms, the action to signal time, and the number of correct, ambiguous, wrong and null identifications.

The rest of the paper is structured as follows. In Section 2 the proposed strategy is described. Next the results of the performance comparison are presented and discussed. A "Conclusion" section closes the work.

2. A Nearest in Control Based Method for Batch Monitoring

A Multivariate Statistical Process Control strategy is devised based on the NICN concept. It has two stages that are performed off-line and on-line respectively.

2.1. Off- line Stage

Normal operation data of batch processes are commonly grouped in a three-way data matrix $\underline{\mathbf{Z}}$ ($I{\times}J{\times}K$), where I stands for the number of batches, J represents the number of variables and K indicates the number of observations of each variable during the batch run. After defining the Normal Operating Condition (NOC) data set, this stage involves the following steps:

a. Each column of $\underline{\mathbf{Z}}$ is centred and scaled to obtain matrix $\underline{\mathbf{X}}$. The mean $\bar{z}_{j,k}$ and the standard deviation $ds_{j,k}$ used to standardized the (j,k)-th column are stored.

b. The inverse of the correlation matrix of each time slide of $\underline{\mathbf{X}}$, $\mathbf{R}_k^{-1}$, is calculated and stored.

c. The T^2 statistic is calculated at each interval k for the members of the reference population. Then, the probability density-function of T^2 corresponding to interval k is estimated using a kernel smoothing technique. Given the significance level of the test, α, the (1- α) percentile of this distribution is selected as the critical value $T_{crit,k}^2$.

Dynamic Time Warping techniques (Ramaker et al. 2003) are successfully applied if the batches included in the NOC data set have different duration or some measurements are missed or delayed. Furthermore matrix inversion-related problems may appear if the number of batches is not sufficient, if the number of variables is quite large or if some of them are linear combinations of others. This problem might be addressed by applying regularization techniques to $\mathbf{R}$ in order to make it non-singular.

2.2. On- line Stage

During batch operation a measurement vector $\mathbf{z}_k$ of dimension J is tested for monitoring purposes at each time interval k. Testing $\mathbf{z}_k$ involves the following steps:

a. The standardized observation vector $\mathbf{x}_k$ is calculated using the mean $\bar{z}_{j,k}$ and the standard deviation $ds_{j,k}$ values corresponding to the j-th variable which are obtained from the NOC.

b. The value of the Hotelling's statistic for $\mathbf{x}_k$, $T_{x,k}^2$ is estimated as follows:

$$T_{x,k}^2 = \mathbf{x}_k^T \mathbf{R}_k^{-1} \mathbf{x}_k \qquad\qquad (1)$$

where $\mathbf{R}_k^{-1}$ is an estimate of the correlation matrix ($\mathbf{\Sigma}$) and is also extracted from the NOC.

c. The values of $T_{x,k}^2$ and $T_{crit,k}^2$ are compared to detect faults. Even if there are no systematic errors in measurements, the Hotelling's statistic value for an observation point may exceed the critical value $T_{crit,k}^2$ for a given significance level. This arises when one or more variables in the current observation vector do not behave as measurements in the normal population do. In this case an anomalous event is detected and next the faulty variables should be identified.

d. Fault identification. The knowledge about how far the faulty observation is from an in control allocation gives us an idea of the minimum distance and direction that need to be explained for an anomalous situation. This information can be obtained finding the nearest neighbour of the observation point that is in statistical control ($\mathbf{x}_{NICN,k}$). For this purpose an optimization problem is formulated whose objective is to minimize a distance measure between $\mathbf{x}_{NICN,k}$ and the measured point $\mathbf{x}_k$, subject to the constraint that the T^2-value for $\mathbf{x}_{NICN,k}$ is equal to or less than the critical one ($T_{crit,k}^2$). If the Mahalanobis distance is selected as the distance measure, the following optimization problem arises

$$Min \quad (\mathbf{x}_k - \mathbf{x}_{NICN,k})^T \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{x}_{NICN,k})$$
$$s.t. \tag{2}$$
$$\mathbf{x}_{NICN,k}^T \mathbf{R}_k^{-1} \mathbf{x}_{NICN,k} \leq T_{crit,k}^2$$

The solutions of the first order optimality conditions for the constrained optimization problem are

$$\mathbf{x}_{NICN,k} = \pm \left(\frac{T_{crit,k}^2}{T_{x,k}^2} \right)^{1/2} \mathbf{x}_k \tag{3}$$

As there exist only two possible solutions, the comparison of the objective function values at both solutions results easier than the evaluation of the second order optimality conditions to decide which one is the corresponding nearest neighbour.

The contribution of each variable to the fact that, T^2 for the current measurement is greater than $T_{crit,k}^2$, can be estimated in several ways using this information. In this work a simple method is implemented and evaluated. Since measurements are standardized, the movement in each direction that should be performed to reach the nearest in statistical control point from the measurement point is used as an estimate of the deviation degree of the variable. Therefore the directions whose changes are greater than a threshold value τ are associated with suspicious variables.

The procedure consists in comparing the absolute difference between x_k^j and $x_{NICN,k}^j$ for each variable j with a threshold value τ. If for a given variable this difference is greater than τ, it is identified as suspicious. Two values of τ are tested. One of them, τ_1, is the mean of the vector of absolute differences between $\mathbf{x}_k$ and $\mathbf{x}_{NICN,k}$. The other one, τ_2, is calculated using the mean plus the standard error of the same vector

$$\tau_1 = \frac{1}{J} \sum_1^J \left| x_k^j - x_{NICN,k}^j \right| \tag{4}$$

$$\tau_2 = \frac{1}{J} \sum_1^J \left| x_k^j - x_{NICN,j}^j \right| + \frac{ds\left(\left| \mathbf{x}_k - \mathbf{x}_{NICN,k} \right| \right)}{\sqrt{J}} \tag{5}$$

3. Application Results

The classic MPCA approach that uses contribution plots in the identification stage and the NICN method are applied for monitoring a non-isothermal semi-batch reactor for methyl-methacrylate emulsion polymerization.

For this purpose a NOC data set composed of 121 batches is obtained by simulation. It is assumed that eight measurements (monomer inlet flowrate Q_0, refrigerant inlet temperature T_{REF0}, jacket temperature T_{REF}, reactor temperature T, surfactant concentration in aqueous phase S_W, liquid holdup h_R, density ρ_R and molecular weight MW) are sampled every 2 minutes for an 80 minute run. The reactor model was developed and implemented in gPROMS code environment (Process System Enterprise, Ltd.) by Alvarez (2009).

Eighty two faults are simulated considering potential process-fault sources. Faults are arranged in eight groups. The first set of faults (1 to 23) comprises perturbations in Q_0. In some cases (1 to 5) Q_0 is fixed at a given value during the whole batch run. The remaining faults in this set (6 to 23) involve step increases or decreases of different sizes that occur at different times during the reactor operation.

The second set of faults (24 to 46) groups several perturbations in the value of T_{REF0}. For faults 24 to 28, T_{REF0} is fixed at a given value during the complete batch run. In contrast, faults in the range 29 to 46 correspond to ramp perturbations in T_{REF0}.

Finally, sets 3 to 8 present both positive and negative deviations in the initial conditions for the liquid holdup, initial surfactant amount, water volume fraction, initiator amount, monomer concentration and reactor temperature respectively.

The performance measures selected for comparison purposes are: the number of false alarms (*FA*), the number of missing alarms (*MA*), the action to signal time (*AST*), and the number of correct (*CI*), ambiguous (*AI*), wrong (*WI*) and null (*NI*) identifications.

The action to signal time accounts for the delay between the fault occurrence and its detection. In this work the *AST* is expressed in terms of the observations' ordinals. In order to obtain a common basis to average *AST* values for both techniques, alarm signals at $k=42$ (one period after finishing the operation) are considered for all missing alarms.

Furthermore the identification is considered correct when only the variable in which the fault was simulated is pointed out by the identification methodology. If this variable and others are indicated as suspicious, a misidentification occurs. A wrong identification arises if the real faulty variable is not identified as suspicious. Finally null identifications correspond to an empty set of faulty variables.

Results are summarized in Fig. 1 which presents the time intervals in which the faults are simulated (k_S) and detected (k_D) for α=0.05, the *AST* and the set of suspicious variables. They are represented using the numbers 2 or 3 depending on the thresholds τ_1 or τ_2 are surpassed. The level of significance of each statistic is selected in such a way that they have the same Average Number of Type I Errors under the null hypothesis.

Figure 1 shows that the last six simulations in group 1, which correspond to small positive and negative magnitude steps in Q_0, are not detected independently of the time interval in which the fault appears.

	#	k_x	k_p	AST_D	Q_0	T_{REF0}	T_{REF}	T	S_W	h_R	ρ_R	MW	k_p	AST_{SPE}	Q_0	T_{REF0}	T_{REF}	T	S_W	h_R	ρ_R	MW	k_p	AST_{T2}	Q_0	T_{REF0}	T_{REF}	T	S_W	h_R	ρ_R	MW
Faults Group 1	1	1	14	13	3							3	22	21	3						3		18	17	3							
	2	1	3	2	3								3	2	3	3			2	2			3	2	3							
	3	1	3	2	3								3	2	3	3	3	3			3	3	3	2	3							
	4	1	3	2	3								3	2	3	3							3	2	3							
	5	1	3	2	3		2			3			3	2	3	3	3	3	3		2		3	2	3							
	6	7	9	2	3				2				9	2	3		2		3			9	2	3								
	7	17	19	2	3						3		19	2	3				2			19	2	3								
	8	27	29	2	3								30	3	3							29	2	3								
	9	7	9	2	3								9	2	3	2	2		3		3	9	2	3								
	10	17	19	2	3								19	2	3							19	2	3								
	11	27	29	2	3								32	5	3				3			29	2	3								
	12	7	9	2	3								9	2	3							9	2	3								
	13	17	21	4	3								16	-1					3			21	4	3								
	14	27																														
	15	7	9	2									9	2	3				2			9	2	3								
	16	17	22	5	3		2		3				22	5	3							22	5	3							2	
	17	27																														
	18	7																														
	19	17																														
	20	27																														
	21	7																														
	22	17																														
	23	27																														
Faults Group 2	24	1	3	2		3			3				26	25			3				2	3	2		3							
	25	1																														
	26	1																														
	27	1	13	12		3		3	2																							
	28	1	3	2		3							15	14		3		3				3	2		3							
	29	1	15	14	3	3																15	14		3		3					
	30	1	19	18	3																	19	18		3							
	31	1	27	26	3								25	24		3					2	22	21		3		3					
	32	1	14	13	3																	14	13		3	3						
	33	1	18	17	3	2																17	16		3							
	34	1	22	21	3	2	3															21	20		3							
	35	1	11	10	3				2	3			19	18		3						9	8	2	3		3					
	36	1	11	10	3					3			19	18		3					2	11	10		3							
	37	1	10	9	3								17	16		3		3				10	9		3						2	
	38	1	11	10	3								15	14		3	3		2	2		11	10		3							
	39	1	12	11	3	2			2				15	14		3		3	2			11	10		3							
	40	1	11	10	3								13	12	3	3						11	10		3							
	41	1	12	11	3								15	14		3		3				11	10		3							
	42	1	11	10	3								26	25	3	3	2					11	10		3							
	43	1	12	11	3								22	21	2	3	3		2	3	2	15	14		3							
	44	1	11	10	3	3							15	14		3	3					11	10	2	3		3					
	45	1	11	10	3								23	22		3	3					11	10		3							
	46	1	14	13	3					3			25	24		3			2	2		14	13		3	2		2				
F. Group 3	47	1	3	2					3	3			5	4		3			3			3	2					3				
	48	1	3	2					3	3			5	4		3			2	3		3	2					3				
	49	1	3	2					3				34	33			3			3	3	3	2					3				
	50	1	3	2					3	3			37	36			3			2		3	2					3				
	51	1	3	2					3				7	6	3	2	2	2		3		3	2					3				
	52	1	3	2					3				3	2	2	3			3	3		3	2					3				
F. Group 4	53	1											3	2					3			3	2					3				
	54	1											3	2	2				3			3	2					3				
	55	1																														
	56	1																														
	57	1																				3	2					3				
	58	1											36	35			3				3	3	2					3				
F. Group 5	59	1	17	16			3	3	3		3	3	7	6							3	15	14				3			2	2
	60	1	20	19			3	3		2		3	13	12				2			3	3	20	19				3	3			
	61	1																														
	62	1											28	27			3			2		28	27			3			2			
	63	1											34	33					3	3		25	24			3	2					
	64	1	24	23		3	3	2					15	14			3				3	24	23			3	2					
F. Group 6	65	1																														
	66	1																														
	67	1																														
	68	1																														
	69	1											24	23	2				3			26	25					2		3		
	70	1											25	24	2				2		2	3										
F. Group 8	77	1	3	2			3	3		2	3		3	2		3	3	3			2	3	2					3				
	78	1	6	5			3						3	2				3	3			3	2					3				
	79	1	9	8	3								3	2	3	3						3	2					3				
	80	1	4	3									3	2			3	3	3			3	2						3			
	81	1	4	3	3		3	3					3	2	2	3	3	3		3		3	2					3				
	82	1	3	2			3	3	3		3	3	3	2		3	3	3				3	2					3	2			

Figure 1. Performance Results

The same happens with faults 14 and 17 although the step sizes (the first positive and the last negative) are considerable higher. This problem arises because faults occur during the last period of the reaction which is less sensitive and has high variability.

For faults 29 to 46, the *ASTs* are, in general, greater than those corresponding to previous faults for all the statistics. This behaviour is motivated by the fact that the simulated perturbations (temperature ramps) are not drastic enough to cause an alarm signal immediately. Faults in groups 5, 6 and 7 are changes in the initial conditions simulated in unmeasured variables. Therefore, they can be detected depending on their effects on measurements. As the identification stage provides no clear indication, operator's knowledge is always required in this sort of situations. It can be seen that many of the missing alarms occur in these groups. In fact, no one of the faults in set 7 was detected. In contrast, sets 3, 4 and 8 involve faults in the initial values of some measured variables, and the faulty variable is properly indicated in most of the cases. Results provided in Figure 1 are summarized in Table 1 for comparative purposes.

Table 1. Summary of performance results

		FA	*MA*	*AST*	*CI*	*AI*	*WI*	*NI*
MPCA	*D*	0	31	21	24	24	1	2
	SPE	1	30	23	4	44	3	0
NICN	T^2	0	25	18	41	15	1	0

The only *FA* obtained for the faulty batches' data set is given by *SPE* statistic for batch 13 at k=16. Table 1 shows that the smallest average value for *AST* and the lowest value of *MA* correspond to the T^2 statistic (NICN's metric). Furthermore it should be noticed that the highest values for *CI* as well as the lowest values for *AI* are achieved using the NICN approach. Regarding *WI*, there are no significant differences.

4. Conclusions

In this work a Statistical Control strategy for batch processes, which is applied in the space of the original variables, is presented. The distance between the observation and its nearest in-control neighbour is used to identify suspicious measurements.
A performance comparison between the NICN technique and MPCA for monitoring a batch polymerization reactor is conducted. Regarding the detection capability, higher values of *MAs* and *ASTs* are obtained when statistics are calculated in terms of latent variables for the tested set of faults. Furthermore, the proposed method shows a rewarding performance to correctly identify the fault source, avoiding ambiguous identifications. Regarding the threshold values, the higher threshold behaves better because it reduces the *AI* without increasing the *NI*.

References

C. Alvarez, 2009, Monitoreo de Procesos Batch. Aplicaciones a Reactores de Polimerización, PhD Thesis, Universidad Nacional del Sur.
C. Alvarez, A. Brandolin, M. Sánchez and L. Puigjaner, 2008, Proceedings of 2008 AIChE Annual Meeting.
P. Nomikos and J. Macgregor, 1994, Monitoring of Batch Process Using Multiway Principal Component Analysis, AIChE Journal, 40, 1361-1375.
J. Westerhuis, S. Gurden, A. Smilde, 2000, Generalized Contribution Plots in Multivariate Statistical Process Monitoring, Chemom. Intell. Lab. Syst., 51, 95-114.
H. Ramaker, E. N. M. Van Sprang, J. A. Westerhuis y A. K. Smilde, 2003, Dynamic Time Warping of Spectroscopic Batch Data, Analytica Chimica Acta, 498, (1-2), 133-153.

1587

Fouling Management in Crude Oil Preheat Trains Through Stream Split Optimization

Joana L. Borges,[a] Eduardo M. Queiroz,[a] Fernando L. P. Pessoa,[a] Fábio S. Liporace,[b] Sérgio G. Oliveira,[b] and André L. H. Costa,[c]

[a]*Federal University of Rio de Janeiro – UFRJ, School of Chemistry - Department of Chemical Engineering, Av. Horácio Macedo, 2030, CT, Bl. E, Cidade Universitária, CEP 21941-909, Rio de Janeiro - RJ, Brazil*
[b] *PETROBRAS/CENPES/PDEDS/Gás Natural/Célula de Otimização, Av. Horácio Macedo, 950 - Cidade Universitária, Rio de Janeiro, 21949-915, Brazil.*
[c]*Rio de Janeiro State University – UERJ, Institute of Chemistry - Department of Industrial Operations and Designs, Rua São Francisco Xavier, 524, Maracanã, CEP 20550-900, Rio de Janeiro – RJ, Brazil*

Abstract

In general, the first main step of petroleum refining consists in the distillation of the crude oil stream. In order to provide adequate fractionation, the crude stream must be fed in the atmospheric distillation column at about 380ºC. Aiming to reduce energy consumption, heat from hot streams of side products and pumparounds is transferred to the crude stream in a heat integration scheme, called crude preheat train. The final heating of the crude stream is executed in a furnace. However, during the operation of the preheat train; the thermal effectiveness of the heat exchangers diminishes due to fouling and, as a consequence, fuel costs increases. The large volumes of crude oil processed and the scenario of crescent energy prices justify the importance of this problem for the oil companies. Seeking to provide a solution to reduce the impact of this problem, this paper presents the exploration of stream splitting in crude preheat trains composed by several parallel branches. In this case, each branch may present different fouling levels, which allows the exploration of different distributions of the stream flow rates along the system, through a proper optimization algorithm. This optimization algorithm searches the set of stream splitters related to the maximization of the final temperature of the crude preheat train. A mathematical model of the preheat train works coupled to the optimization method. An important focus of this paper is to explore the introduction of constraints in order to guarantee feasible operating solutions, i.e., the optimum solution must attend different operational aspects related to bounds on fluid flow velocities and heat exchanger capacities. The performance of the proposed approach is illustrated through a typical example of a petroleum refinery.

Keywords: fouling, heat exchangers, oil refining.

1. Introduction

Conversion of crude petroleum into more valuable products involves many operations, such as, distillation, catalytic cracking, hydrotreating, delayed coking, etc. The first step frequently involves an atmospheric distillation column, sometimes preceded by a pre-flash column.

Crude petroleum must be pre-warmed to a satisfactory temperature before it is fed into those columns. Usually there is a huge consumption of energy at this stage. Therefore, aiming the reduction of fuel costs (with consequently decrease of gas emissions), energy integration of side-product streams and pumparounds with the crude stream is promoted. The final heating step is conducted in a furnace. Such structure is denominated preheat train.

An important operating problem of this kind of structure corresponds to the fouling of thermal surfaces. This phenomenon diminishes the thermal effectiveness and increase the furnace load. Recently, the literature presented several papers concerning the optimization of cleaning actions (Smaïli *et al.*, 2001; Lavaja and Bagajewicz, 2004).

In many preheat trains, the heat exchangers are organized in multiple parallel branches which can suffer fouling distinctly, because of different previous cleaning records or operating policies. Oliveira Filho *et al.* (2007) explored the optimization of stream splitters in multi-branch preheat trains aiming to maximize energy recovery and minimize operational costs.

In this context, this paper studies the application of the proposed technique in a real refinery, analyzing important questions in order to guarantee an actual operational solution: complexity of the control loops, maximum capacity of the coolers, impact of the variation in the pumparounds on the distillation performance and modification of fluid velocities in the fouling behavior. Economic results are also evaluated, seeking to measure the potentiality of the methodology.

2. Methodology

In a multi-branch preheat train, the final temperature depends on the flow distribution. Thus, the employed methodology is based on an association of a simulation algorithm to an optimization procedure. For each set of split fractions along the branches, a final crude temperature can be evaluated. The optimization seeks to find the best set of split fractions in order to reach the highest temperature (associated to the lowest fuel cost). The proposed scheme is illustrated in Figure 1.
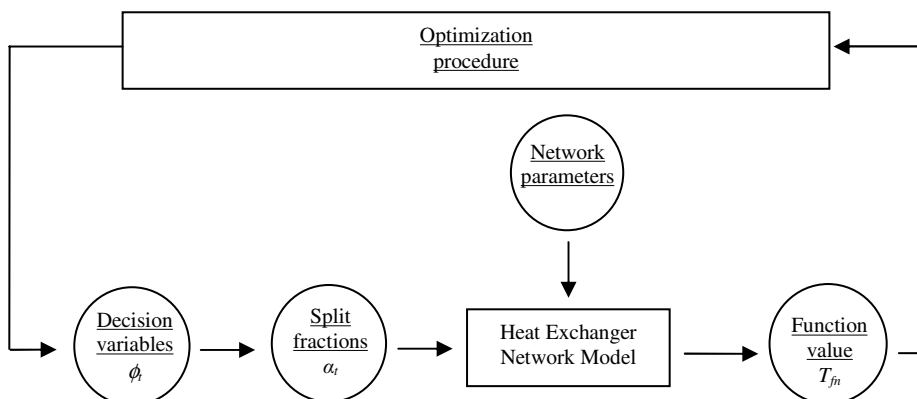
Figure 1 - Objective function evaluation scheme.

2.1. Feasibility issues

The feasibility of the application of the proposed approach in a real refinery involves the solution of some limitations of the existent hardware.

2.1.1. Complexity of the control loops

The maximum extension of the degrees of freedom would involves the insertion of control valves for all splitters. However, it can imply an unnecessary capital cost with process instrumentation. This analysis demands the investigation of different sets of control arrangements, comparing the energy reduction gain with the complexity of each control loop set.

2.1.2. Maximum capacity of the coolers

The maximization of the crude stream temperature implies, in certain situations, an increase of the inlet product stream temperature in a final cooler. This effect can influence the product storage (with a possible disproportional increase of cooling water demand). This aspect is handled by the insertion of constraints to avoid overcharging of the external heat exchangers.

2.1.3. Impact of the variation in the pumparounds on the distillation performance

Another possible disturbance that can be caused by the optimization procedure consists in the pumparound operation. In order to avoid a misbalance of the heat load along the distillation column, such problem is also solved by the insertion of bounds on these heat exchangers.

2.1.4. Modification of fluid velocities in the fouling behavior

Recently, Ishiyama *et al.* (2008) argued that the difference in flow rates of the branches, caused by the splitters misbalance, would accelerate the fouling process instead of contributing for the energy recovery. In fact, the fouling rate decreases with increasing flow velocity, as a consequence, the streams splitters optimization could generate results where a branch would receive a too low crude velocity that the fouling would consistently raise and the energy recovery of the system would be diminished. This aspect can be contemplated in the optimization through the imposition of velocities bounds in heat exchangers, thus avoiding very low velocities.

3. Example

The example describes a real network of a Brazilian refinery. The complete refinery crude preheat train presents 35 heat exchangers organized in five branches. The process flowsheet can be found in Figure 2. The first studied case (reference) corresponds to a simulation of the preheat train with all heat exchangers without fouling. The second case (base) is similar to the reference case, but considering the operational fouling of the refinery (at a specific day).

3.1. Analysis of the control loops

The investigation of the control loops is conducted by the examination of nine splitter schemes, described in Table 1, without constraints. For each scheme, the optimization was run and the final temperatures and the corresponding economic results are presented in Tables 2 and 3, respectively.

J.L. Borges et al.
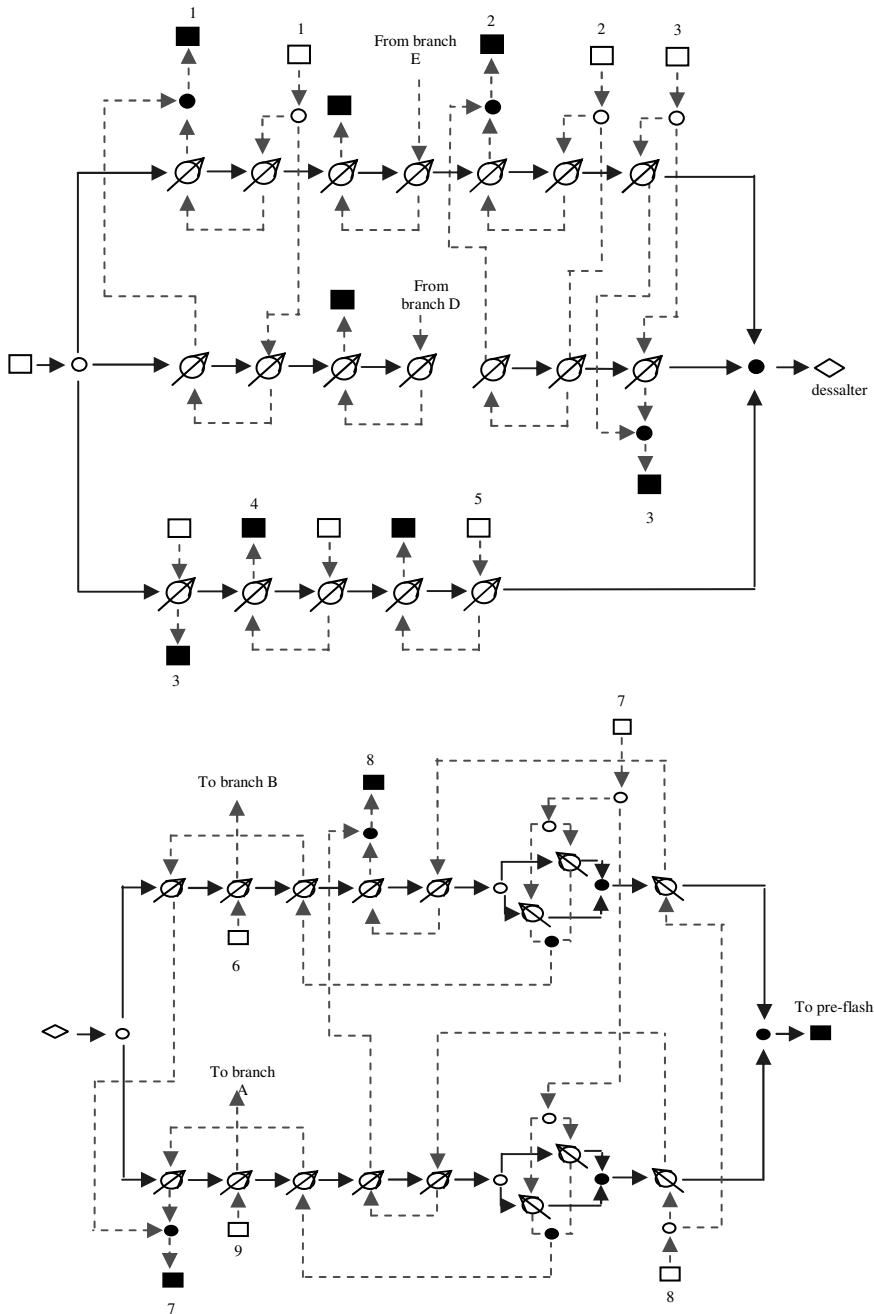
Figure 2 - Heat exchanger network flowsheet upstream and downstream the desalter -
Cold streams: continuous lines - Hot streams: dashed lines - Supply nodes: white
squares - Demand nodes: black squares - Splitters: white circles - Mixers: black circles

Table 1 – Optimized splitters on analyzed cases

Cases		1	2	3	4	5	6	7	8	9
Upstream the	**Hot**	Yes	-	Yes	-	Yes	-	Yes	Yes	Yes
desalter	**Cold**	Yes	Yes	Yes	Yes	Yes	-	-	Yes	-
Downstream	**Hot**	Yes	-	-	Yes	-	Yes	Yes	Yes	Yes
the desalter	**Cold**	Yes	Yes	Yes	Yes	-	Yes	-	-	Yes

Table 2 - Network final temperatures (ºC)

Reference	Base	1	2	3	4
273.94	265.15	266.22	265.84	265.85	266.22
5	**6**	**7**	**8**	**9**	
265.46	265.92	265.80	266.09	265.93	

Table 3 – Reduction of fouling costs (MMUS$/year)

Reference	Base	1	2	3	4	5	6	7	8	9
-	-	0.17	0.11	0.11	0.17	0.05	0.12	0.10	0.15	0.12

The temperature gain ranges from 3.43% to 12.12% in relation to the total fuel costs due to fouling, evaluated by $(T_{reference} - T_{case}) / (T_{reference} - T_{base})$. The best results involve the optimization of all splitters (case 1), however similar temperature increases are obtained by simpler configurations (case 4 and 8). In terms of savings, it can be seen that even a 1 °C reduction, such as obtained, represents a great economy on operational costs of refinery due to fouling.

3.2. Feasibility constraints

The constraints discussed in the previous section were introduced in the optimization procedure by the insertion of penalty factors. The control loop scheme employed was case 1 (all splitters optimization).

Temperature constraints are the same in all cases: (i) Product streams: the limits were calculated based on a 20% upper variation on the heat transfer rate of the cooler needed to storage of products; (ii) Pumparounds streams: the limits appeared based on a variation on the heat load of the heat exchangers (upper and lower bounds). The constraints on the heat exchanger velocities were analyzed by several criteria, presented in Table 4.

Table 4 – Constraints in the heat exchanger velocities

Case	Lower limit	Upper limit
10	None	None
11	0.9 m/s in all heat exchangers	2.4 m/s in all heat exchangers
12	- 20% variation of base case	+ 20% variation of base case
13	Smallest lower limit of 11 and 12	Highest upper limit of 11 and 12
14	0.9 m/s in all heat exchangers	None
15	- 20% variation of base case	None
16	Smallest lower limit of 11 and 12	None

The final crude oil temperature in all cases is shown in Table 5. Attending the constraints implies just a small decrease of the possible temperature gain.

Table 5 – Optimized splitters on analyzed cases

Cases	10	11	12	13	14	15	16	17
Temperature (^{o}C)	266.22	266.21	266.19	266.09	266.17	266.20	266.17	266.17
Solution feasibility	-	Yes	No	Yes	Yes	No	Yes	Yes

It is important to mention that none of the streams had its temperatures violated with the optimization. Case 10 results were very similar to the optimizations without constraints once the optimized case originally was already inside the limits. Velocity constraints based on minimum velocity of 0.9 m/s and maximum velocity of 2.4 m/s (case 11) were not attended since a feasible case could not be found. However, it is important to mention that all heat exchangers of the refinery analyzed in operational conditions already violated those velocity bounds. That is the reason why it was proposed the establishment of a 20% velocity variation in relation to the base case.

4. Conclusion

Results show that optimization of hot and cold streams splitters can have a great impact on operational costs of refineries and should be further investigated. The careful analysis of operating aspects allowed the identification of realistic solutions for the fouling management in multi-branches preheat trains.

Acknowledgments

References

E. M. Ishiyama, W. R. Paterson, D. I. Wilson, 2008, Thermo-hydraulic channelling in parallel heat exchangers subject to fouling, Chemical Engineering Science, 63, 3400-3410.

J.H. Lavaja, M.J. Bagajewicz, 2004, On a new MILP model for the planning of heat-exchanger network cleaning, *Ind. Eng. Chem. Res.*,43, 3924-3938.

F. S. Liporace, S. G. Oliveira, 2005, Real time fouling diagnosis and heat exchanger performance, in: Proceedings of 6th International Conference on Heat Exchanger Fouling and Cleaning – Challenges and Opportunities, Kloster Irsee, Germany, 267-277.

L.O. Oliveira Filho, F. S. Liporace, E. M. Queiroz, A. L. H. Costa, 2007, Investigation of an alternative operationg procedure for fouling management in refinery crude preheat trains, *Applied Thermal Engineering*, submitted for publication, 2007.

F. Smaïli, V.S. Vassiliadis, D.I. Wilson., 2001, Mitigation of fouling in refinery heat exchanger networks by optimal management of cleaning, *Energy & Fuels*. 15, 1038-1056.

1593

Optimal Sensor Location for Chemical Process Accounting the Best Control Configuration

David Zumoffen[a,b] and Marta Basualdo[a,b]

[a]*Grupo de Informática Aplicada a Ingeniería de Procesos (GIAIP),
CIFASIS-CONICET-UNR. 27 de Febrero 210 bis, S2000EZP Rosario, Argentina.
E-mail address: zumoffen, basualdo@cifasis-conicet.gov.ar*
[b]*Universidad Tecnológica Nacional (UTN), FRRO. Rosario, Argentina.*

Abstract

In this work a new methodology for solving simultaneously the problems of optimal sensor location and control structure selection for large scale chemical processes is presented. Here, it is considered the need of guaranteing the best plant-wide control structure before answering about which is the best sensor net able for achieving that objective. In this work it is demonstrated the importance of answer both questions as an integrated problem because of the strong impact in the initial investment and the future controlled process performance. Most of the previous works in this area analyze these problems as separated subjects. Here, genetic algorithms (GA) are used because they represent a valuable tool for support the decisions about the sensor placement, possible pairing of input output variables among a great number of combinations, since the interaction effect point of view. It allows to avoid the expert knowledge as decision criteria for pairing selection. The preliminary study is done on a simplified plant model obtained by subspace identification techniques (4sid). The final testing is performed on the rigorous dynamic model with the obtained plant-wide structure where the controllers tuning is performed through the internal model control (IMC) theory. The well-known case of the Tennessee Eastman (TE) benchmark is adopted for testing the methodology described here and compared with other strategies.

Keywords: sensor location, genetic algorithm, plant-wide control, RGA.

1. Introduction

In this work a new methodology for optimal sensor location accounting the control structure explicitly for large scale chemical processes is presented. It was found that from the optimal sensor location point of view the problem to be solved considers the sensor network design taking into account investment costs and observability topics (Musulin et al., 2005; Kadu et al., 2008; Singh and Hahn, 2005). Typical tools such as Kalman filtering, GA and pareto graphics were used on those papers. All of these works considered an open loop process or an existing controlled process. A similar situation occurs in the control structure selection area, focused in process interaction measures without considering any integration with the optimal sensor location problem. As can be seen, most of the previous works did not analyze this problem simultaneously, which is one of the new results introduced here. The TE benchmark control problem with the requirements imposed by Downs and Vogel (1993) is used for testing the proposed methodology. The steps for its implementation consist on: firstly performs the process stabilization which is carried out by three level control loops. Secondly, the subspace identification techniques (4sid) are applied for obtaining a linear

state-space model accounting those more critical requirements such as production rate, G/H ratio and B composition in purge. The problem to solve is the proper measurements selection with the minimum number of sensors located on optimal positions from the control point of view. In this context, it is proposed to find a control structure with smaller interaction chosen among 495 variables combination. Usually, the literature suggests heuristic criteria in order to reduce the number of combinations. The objective here is to develop a systematic approach that rigourously drives to the plant-wide control structure without heuristic considerations. Since the system is non square, the main ideas given by Chang and Yu (1990) were accounted, even though they could applied their technique only on plants of low dimensionality. In this paper the extension of Chang and Yu (1990) methodology is done with the help of GA. Finally, the IMC theory in steady-state for full controller design and the previously identified linear model support the decision of the outputs selection (to make the process square). As a result an interesting improvement can be achieved for both, control performance and hardware requirements, compared with the results given by McAvoy and Ye (1994). The most prominent simulation results are included according to some of the tests imposed for the TE. In addition, as future work is considered to analyze the advantages obtained when the expert system for diagnosis and fault detection for complex and large chemical process (Zumoffen and Basualdo, 2008a,b) are developed on the bases of the approach proposed here.

2. Case study: Tennessee Eastman process

Essentially, the TE plant generates two products from four reactants and has five principal units of operation: the reactor, the product condenser, a vapor-liquid separator, a compressor and a product stripper. The stripper underflow contains the key components G and H. In this work, it is assumed that the plant is operating under conditions considered as *base case*, that is G/H mass ratio of 50/50 and a production rate of 7038 kgG/h and 7038 kgH/h. The plant dynamic model used here was taken from Ricker's control department of the Washington University (`http://depts.washington.edu/control`).

3. Initial considerations

The TE open loop response is unstable, so, the first step is to stabilize the plant through three level controllers at the reactor, separator and the stripper units respectively. Then, the remaining variables to be considered are: 9 inputs and 12 outputs which are detailed at Table 1. In this work the nomenclature used is the same that the suggested one by Downs and Vogel (1993). The system identification (SI) procedure is performed to obtain a linear model in discrete state-space format. The SI begins with the plant inputs excitation using an uniformly distributed random magnitude between $\pm 2\%$ around the base case operation point and a period of 100 hr.. The recorded data of this process are collected using a sampling time of 5 min..

4. Optimal sensor location using GA

According to the TE dimensionality the linear model has $m = 12$ potential measurements and $n = 8$ potential manipulated variables because $XMV(12)$ has no impact on any of the considered output. Therefore, for detecting the less interactive pairing, the development given by Chang and Yu (1990) is accounted. They proposed a pairing selection supported by RGA for non square systems. It involves consider the plant model as $G = [G_s, G_r]^T$ displayed in the Fig. 1 of $m \times n$ dimension, which is partitioned into G_s of $n \times n$ and a G_r

Table 1: Input-output variables and base case values

Inputs	Variable	Base values	Outputs	Variable	Base values
D feed	$XMV(1)$	63.053 %	Recycle flow	$XME(5)$	26.902 kscmh
A feed	$XMV(3)$	24.644 %	Reactor feed	$XME(6)$	42.339 kscmh
A and C feed	$XMV(4)$	61.302 %	Reactor pressure.	$XME(7)$	2705.0 kPa gauge
Comp. rec. valve	$XMV(5)$	22.210 %	Reactor temp.	$XME(9)$	120.40 °C
Purge valve	$XMV(6)$	40.064 %	Separator temp.	$XME(11)$	80.109 °C
Stripper steam valve	$XMV(9)$	47.446 %	Separator pressure	$XME(13)$	2633.7 kPa gauge
RCWO t. set point	$XME(21)sp$	94.599 °C	Stripper pressure	$XME(16)$	3102.2 kPa gauge
CCW flow	$XMV(11)$	18.114 %	Production rate	$XME(17)$	22.949 m³/h
Agitator speed	$XMV(12)$	250 rpm	Stripper temp.	$XME(18)$	65.731 °C
			Compressor work	$XME(20)$	341.43 kW
			B comp. in purge	$XME(30)$	13.823 mol%
			G/H ratio	$XME_{G/H}$	1.2257
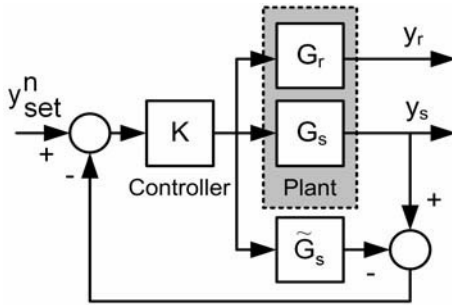
Figure 1: IMC for non-square plant
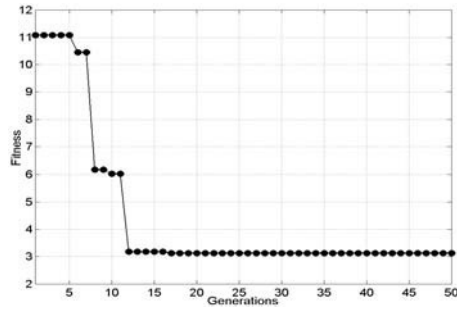
Figure 2: Fitness function

of $(m-n) \times n$. Then,

$$y = \begin{bmatrix} y_s \\ y_r \end{bmatrix} = \begin{bmatrix} G_s \\ G_r \end{bmatrix} u \tag{1}$$

where y_s and y_r are $n \times 1$ and $(m-n) \times 1$ respectively. Considering the IMC theory, a square controller can be designed as $K = \tilde{G}_s^{-1}$, so $u = G_s^{-1} y_{set}$, in addition, perfect plant-model $G_s = \tilde{G}_s$ is assumed and output disturbances are neglected. The closed loop error e is defined as,

$$e = y_{set}^m - y = y_{set}^m - GG_s^{-1} y_{set}^n \tag{2}$$

where $y_{set}^m = [y_{set}^n, y_{set}^{m-n}]^T$, $y_{set}^n = [y_{set}^1, \ldots, y_{set}^n]^T$. Considering $y_{set}^{m-n} = [0, \ldots, 0]^T$ then,

$$e = \left[\begin{bmatrix} I_{n \times n} \\ 0_{(m-n) \times n} \end{bmatrix} - GG_s^{-1} \right] y_{set}^n = A y_{set}^n, \quad \text{with} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \tag{3}$$

being e $m \times 1$ and considering all the system outputs. Then, eq. 3 can be calculated for the rest of set point changes so as $e(i) = A y_{set}^n(i)$. Here $e(i)$ makes reference to the error on all variables when an unitary set point change in the variable i has been occurred, being $y_{set}^n(i)$ a vector with unit value in the i entry and zero for the rest of the elements. The SSE index computes the sum of the square errors for the overall plant variables accounting each set point variation, that is:

$$SSE = \sum_{i=1}^{n} ||e(i)||_2^2 = \sum_{i=1}^{n} ||A y_{set}^n(i)||_2^2 = \text{tr}\left(A^T A\right) \tag{4}$$

The control objective is to minimize the SSE. Clearly a particular selection of G_s affects the *SSE* performance. Additionally, according to the number of involved variables , the combinatorial problem presents $m!/(n!(m-n)!) = 495$ possible solutions. The optimal solution is found by using the genetic algorithm (GA) approach (Chipperfield et al., 1994). It is a stochastic global search method that mimics the metaphor of natural biological evolution. In this case, the chromosome length can be associated to the possible sensor set dimension and the sensor network structure with the chromosome genes. Thus, $I_j = [g_1, g_2, \ldots, g_{N_c}]$ is the chromosome, where $j = 1, \ldots, N_i$ and N_i the initial population dimension. The genes parameters g_i, with $i = 1, \ldots, N_c$, belong to a binary code $g_i = \{1, 0\}$ indicating the presence or not of a sensor in the i location, and N_c the potential sensor network length. The problem to be solved can be represented as

$$\min_{I_j} \left[SSE(I_j) \right] \quad \text{subject to: } ||I_j|| = n \tag{5}$$

where the condition $||I_j|| = n$ guarantees the square selection of the sub-process G_s. Solving eq. 5 by GA the optimal sensor location, I_{op}, can be obtained. In the Fig. 2 can be observed the fitness criteria profile for the best individual selected in less than 50 generations. The sensors net for the optimal solution I_{op} is displayed at Table 2, in addition a suboptimal I_{sop} one is also included . The zero values in these chromosomes represent no sensors in these locations, and the opposite for the unitary values. The optimal solution generates the following optimal sensor network structure: *recycle flow, separator temperature, separator pressure, production rate, stripper temperature, compressor work, B composition in purge and G/H ratio.*

5. Control structure

In this section the control structure is defined by using RGA analysis (Bristol, 1966) carried out for the optimal solution. Computing the RGA from Table 2 results can be obtained the relative gain matrices displayed at Table 3 for optimal Λ_{op} case, where $\Lambda_{op} = G_s(I_{op}) \otimes (G_s(I_{op})^{-1})^T$. The best RGA presents lower interaction between the process variables than the suboptimal one. The final control topology is displayed at Fig. 3.

6. Results

The optimal sensor location and control structure selection presented in previous sections are compared and analyzed. By one side it is evaluated if the process operation conditions stated by Downs and Vogel (1993) are met. By other side both control performance and hardware requirements are compared with the classical control structure presented in previous work (McAvoy and Ye, 1994). The TE process presents four possible set point modifications (production rate, product mix, reactor pressure and B purge composition) from its base case value and different possibilities to analyze disturbance effects (IDV1 to IDV20). Due to space limitation only some results are displayed here. The Fig. 4 considers the IDV1 disturbance effects on both control structures and displays the reactor pressure profile. The simulations are done once the controllers tuning, based on the IMC theory (Rivera,

Table 2: Different solutions from GA

Individual	Sensor network												Fitness
I_{op}	1	0	0	0	1	1	0	1	1	1	1	1	3.1280
I_{sop}	0	0	1	1	1	0	0	1	1	1	1	1	11.0751

Table 3: Optimal solution RGA

I_{op}	$XMV(1)$	$XMV(3)$	$XMV(4)$	$XMV(5)$	$XMV(6)$	$XMV(9)$	$XME(21)sp$	$XMV(11)$
$XME(5)$	0	0	0	0	0.17	0.02	1.38	0.331
$XME(11)$	0	0	0.10	0.47	0.31	0	1.45	0.61
$XME(13)$	0.19	1.07	0.11	0.19	0	0	0	0.18
$XME(17)$	0	0.10	1.07	0.03	0.01	0.03	0	0
$XME(18)$	0.06	0.35	0	0	0	1.92	0	0
$XME(20)$	0.04	0.15	0	1.17	0.13	0.03	0	0.02
$XME(30)$	0.08	0.03	0	0	0.68	0.02	0.05	0.17
$XME_{G/H}$	1.05	0.07	0	0	0	0	0	0.01
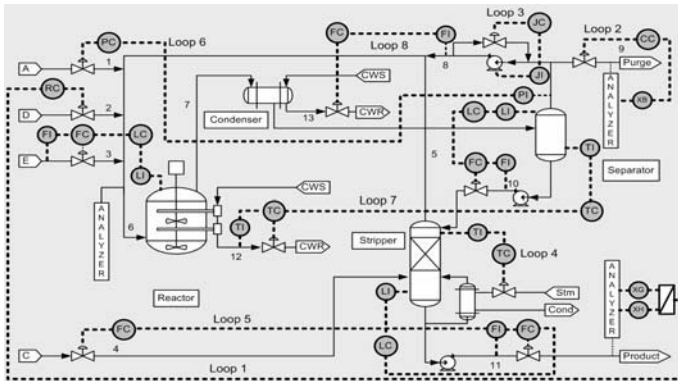
Figure 3: Final control structure

2007) implemented with the dynamic linear model identified previously. All the controllers implemented in this work are discrete PI versions obtained from the continuos case. The control policy obtained from the optimal sensor network clearly shows the good regulatory properties in this case. By other hand, the control structure for the suboptimal sensor network presents problems for rejecting the disturbance. An incorrect pressure control loop produces the plant shut down. The Fig. 5 displays the reactor pressure time evolutions when a production rate change occurs. The suboptimal control structure cannot maintain the correct operation of the process and the shut down (labelled with sd) takes place due to reactor pressure violation. The alternative control structure proposed here fulfills with the specifications showing a similar performance as that presented in McAvoy and Ye (1994). The advantage of the approach presented here is mainly given by a lower number of sensors needed as can be observed at Table 4.

7. Conclusions

In this work has been presented an alternative optimal sensor location method integrated to control structure design for large-scale processes. The final result displays a systematic approach that allows arriving to an efficient decentralized control structure, implemented with the minimum number of measurements. Additionally, this methodology is practically independent of any heuristic information. The simulation results obtained here has been compared with previous works showing clear advantages in performance and investments. As future work, it will be analyzed which is the potentiality for improving expert systems design for fault diagnosis in large chemical plants.
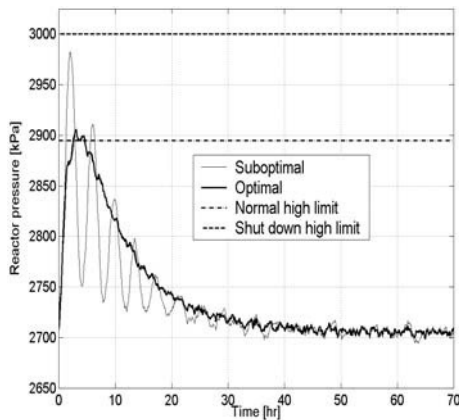
Figure 4: IDV1 disturbance
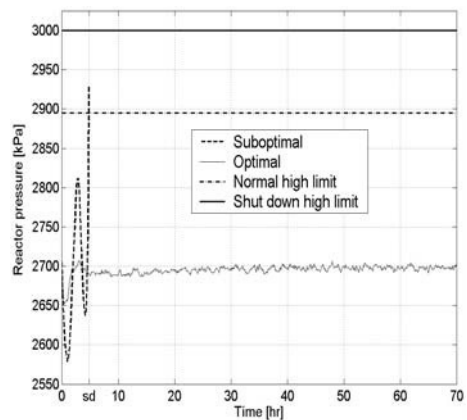
Figure 5: Production rate change

Table 4: Hardware requirements

	McAvoy and Ye (1994)	Ricker (1996)	Larsson et al. (2001)	Proposed here
Measurements	22	16	22	15
Control loops	22	19	17	15

References

E. Bristol, 1966. On a new measure of interaction for multivariable process control. *IEEE Transactions on Automatic Control*.

J. Chang and C. Yu, 1990. The relative gain for non-square multivariable systems. *Chemical Engineering Science*. *45,* 1309-1323.

A. Chipperfield, P. Fleming, H. Pohlheim and C. Fonseca, 1994. Genetic algorithm toolbox. For use with Matlab. *University of Sheffield. Department of automatic control and systems engineering.* `http://www.shef.ac.uk/acse/research/ecrg`.

J. Downs and E. Vogel, 1993. A plant-wide industrial process control problem. *Computers Chem. Engng.*. *17,* 245-255.

S. Kadu, M. Bhushan and R. Gudi, 2008. Optimal sensor network design for multirate system. *Journal of Process Control*. *18,* 594-609.

T. Larsson, K. Hestetun, E. Hovland and S. Skogestad, 2001. Self-optimizing control of a large-scale plant: The Tennessee Eastman process. *Ind. Eng. Chem. Res.*. *40,* 4889–4901.

T. McAvoy and N. Ye, 1994. Base control for the Tennessee Eastman process. *Computers Chem. Engng.*. *18,* 383-413.

E. Musulin, C. Benqlilou, M. Bagajewicz and L. Puigjaner, 2005. Instrumentation design based on optimal Kalman filtering. *Journal of Process Control*. *15,* 629-638.

N. Ricker, 1996. Decentralized control of the tennessee eastman challenge process. *Journal of Process Control*. *6,* 205-221.

D. Rivera, 2007. Una metodología para la identificacion integrada con el diseño de controladores IMC-PID. *RIAI (Rev. Iberoamericana de Autom. e Inf. Ind.)*. *4,* 5-18.

A. Singh and J. Hahn, 2005. Determining optimal sensor location for state and parameter estimation for stable nonlinear systems. *Ind. Eng. Chem. Res.*. *24,* 5645-5659.

D. Zumoffen and M. Basualdo, 2008a. Improvements in fault tolerance characteristics for large chemical plants. Part I: Waste water treatment plant with decentralized control. *Ind. Eng. Chem. Res.*. *47,* 5464-5481.

D. Zumoffen and M. Basualdo, 2008b. Improvements in fault tolerance characteristics for large chemical plants. Part II:Pulp mill process with model predictive control. *Ind. Eng. Chem. Res.*. *47,* 5482-5500.

A New Systematic Approach to Find Plantwide Control Structures

Gonzalo Molina[a,c], David Zumoffen[a,b] and Marta Basualdo[a,b]

[a]*GIAIP – Centro Franco-Argentino de Ciencias de la Información y de Sistemas (CIFASIS-CONICET-UNR-UPCAM III). 27 de Febrero 210 bis, S2000EZP Rosario, Argentina*
[b]*Universidad Tecnológica Nacional (UTN) – FRRo. Rosario, Argentina.*
[c]*FCEIyA-Universidad Nacional de Rosario (UNR). Rosario, Argentina.*

Abstract

A new systematic approach for addressing the problem of the plant-wide control structure, supported by steady-state information is presented. It is tested in a reactor/separator with recycle plant. Several authors have presented different control alternatives for this kind of plant. However, most of them mainly focused on the optimum energy consumption or in the regulator problem only. In both cases the decision of the final control structure was adopted based on several heuristic concepts. In this work both objectives are considered avoiding any heuristic concepts. The approach consists meanly on three steps: optimization, stabilization, and final pairing between manipulated and controlled variables focused on rejecting the most critical disturbances. In addition, the reactor design is subject to ensure quality product specification, cost investment together with the overall plant controllability. Finally, the dynamic model of the plant is used for testing the potentiality of the proposed control structure. Hence, a multivariable tuning procedure, based on internal model control (IMC), is performed for evaluating the dynamic closed loop responses under the worst disturbances scenarios.

Keywords: plant-wide control, optimum energy consumption, disturbance rejection, controllability.

1. Introduction

A large number of contributions on plant-wide control structure synthesis can be classified into the process-based experience (engineering judgment) and mathematical-based criteria. Among the most notable contributors on the first category are McAvoy (1998) and Luyben and co-workers (Belanger and Luyben, 1996, 1997). This methodology is systematic in nature and addresses many of the major issues involved in the plant-wide control problem, such as the effects of recycles and energy integration. A nine step approach is developed by Luyben and co-workers based on the process experience of the group. The main result from this philosophy is an alternative way to achieve a decentralized plant-wide control structure. The second category relies on a rigorous mathematical framework of dynamic theory, constrained optimization and systems analysis. Skogestad and co-workers (Skogestad (2000), Larsson (2000), Alstad and Skogestad (2007)) obtained interesting results using these tools. They focused on the selection of self-optimizing control variables that give the smallest loss in profit. The mathematical approaches given by these authors, although rigorous, are often very difficult to formulate for large scale systems, in many cases they turn computationally very intensive. In addition, the model assumptions can affect the convergence and solution. Hence, in this work a kind of consensus between both categories mentioned above is presented. In order to achieve this purpose some useful modifications are proposed to generate a new systematic approach able to deal with the most critical problems such as size, complexity, and multiple objectives commonly found in the majority of chemical plants.

A representative example in the plant-wide control community is studied, which is very useful for obtaining deep knowledge and comparing with other previous results. In the next section a

brief description of the study case is given. In section 3 the complete approach is detailed step by step. Particularly in the third step the extension of the null space criteria is introduced. It allows to perform the variable pairings selection with less interaction effects. In section 4 the closed loop dynamic responses are shown. An acceptable controllers tuning are given for achieving a good regulation under the most critical disturbances. At this point, the reactor design can be also evaluated and modified if it is necessary in case that the product quality,cost investment and controllability aspects do not meet specification.

2. Case Study Description

The studied plant consists on a reactor and a distillation column interconnected whose data has been taken from Wu. and Yu (1996). It is assumed that an irreversible first order reaction $\mathbb{A} \to \mathbb{B}$ occurs in the continuous stirred tank reactor (CSTR) being the reaction rate k a function of the temperature given by Arrhenius expression $k = k_0.e^{-E/R.T_R}$.

This is an exothermic reaction. The reaction rate is considered constant ($k \approx 0.340843$ hr^{-1}). Some of the reactant $\mathbb{A}$ is consumed in the CSTR and the effluent of the reactor, a mixture of $\mathbb{A}$ and $\mathbb{B}$, is fed into the tray 12 of the 20-tray distillation column. The product $\mathbb{B}$ is obtained from the bottom of the column, while the distillate of the column is recycled to the CSTR. Constant relative volatility $\alpha = 2$ between $\mathbb{A}$ and $\mathbb{B}$ is assumed in the column model. The plant is shown in Fig. 1. The final control structure is also included in this figure. It will be explained in the following sections.

3. The proposed systematic approach

In this section the three main steps of the approach are described.

3.1. Process optimization

Table 1: Case Study Parameters

Parameters, Variables	Base Case	Optimum	Units
Reactor			
Feed Flow Rate (F_0)	460	460	(lb mol/hr)
Feed Flow Composition (z_0)	0.9	0.9	
Recycle Flow Rate (D)	500.15	651.89	(lb mol/hr)
Residence Time	2.5	2.7	(hr)
Column			
Column Feed Flow Rate(F)	960.15	1111.89	(lb mol/hr)
Column Feed Composition (z)	0.5	0.475	(lb mol/hr)
Reflux Flow Rate (L)	1100	862.28	(lb mol/hr)
Distillate Flow Rate (D)	500.15	651.89	(lb mol/hr)
Product Flow Rate (B)	460	460	(lb mol)
Product Composition (x_B)	0.0105	0.0105	
Reboiler Hold Up (N_{reb})	275	275	(lb mol)
Condenser Hold Up (N_{cond})	185	185	(lb mol)
Vapor Boil-Up (VB)	1600	1514	(lb mol/hr)

In this stage a steady state optimization is done in order to minimize the vapor boil up VB of the plant, which is considered the objective function J (Larsson, 2000). Hence, the problem to be solved is $\min_{x,u} J(x,u,d)$ subject to the constraints $f(x,u,d) = 0$, $g(x,u,d) \leq 0$, $y = f_y(x,u,d)$; where $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$ and $y \in \mathbb{R}^{n_y}$ are the states, inputs and measurements, respectively; f is the set of equality constraints corresponding to the model equation; g is the set of inequality constraints that limit the operation and n_x, n_y and n_u are the number of states, measurements and manipulated variables respectively.

This optimization problem is solved by using a steady-state model created in GAMS (software for modeling and optimization). The assumed constraints are: the product quality $x_B \leq 0.0105$, the reactor hold-up $V_R \leq 2400$, and flow rates can be at least two times of the initial value (see Wu. and Yu, 1996). In table 1 can be seen both, the obtained optimal magnitudes and base case values where the disturbances z_0 and F_0 are kept constant. From this step it is straight forward the variables which need to be controlled (Skogestad, 2000)are:

- levels without steady state effect (N_{cond}; N_{reb})

- variables that are active constrain in the optimum (x_B, V_R)

3.2. Plant stabilization

The second step proposes the plant stabilization, which drives to a typical inventory control (McAvoy, 1998). Then most suitable paring between the levels as controlled variables with the manipulated variable is performed looking for the less interactive ones supported by the RGA. Since in all cases they are unstable controlled variables (pure integrator function), the RGA calculations were done considering derivative output with respect to time.

In this case only two possible level control structures are possible. The pairings are: the reboiler level N_{reb} with B and bottom level V_R with F. The control of N_{cond} can be performed with D or with L. In this study, the first one was chosen because changes in D produce less impacts on the reactor level because of the relative size between N_{cond} and V_R. The final inventory control structure for this process is shown in Fig. 1 with grey background.
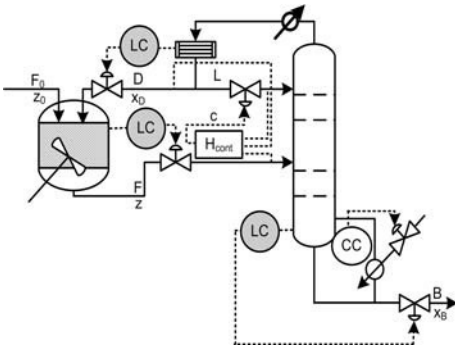
Figure 1: Final Control Structure

Figure 2: Proposed Procedure

Table 2: Measurements Combinations

Nº	Variables	H	RGA
1	$[D,F,L]$	-0.1920 -0.2030 0.7359	0.997
2	$[D,F,z]$	-0.0004 -0.0000 1.0000	0.995
4	$[D,L,z]$	-0.0004 0.0001 1.0000	0.995
7	$[F,L,z]$	-0.0009 0.0002 1.0000	0.995
11	$[B,L,z]$	0.0008 0.0003 1.0000	0.995

3.3. Pairing Variables Selection for Regulatory Control and Economic Optimum

In this stage the main objective is to keep the plant near the optimal steady-state operation without re-optimizing the process when disturbances affect the plant. This is done based on the main concepts of self-optimizing control and null space method introduced by Alstad and Skogestad (2007). It consists on finding the controlled variables c as a linear combination of measurements y. The requirement is to have at least as many measurements as the unconstrained degrees of freedom, including disturbances. This technique was successfully used for controlling a Petlyuk distillation through a choice of a proper combination of temperature measurements. However, for more complex cases, such as the plant studied here, the selected combination of variables was strongly interacted. Because of this, it is proposed performing the selection focused on those combinations that provide a best decentralized control structure supported by RGA calculations.

The linear relationship between c and y is given by $c = Hy$, where H calculation is done through the optimal sensitivity matrix F whose elements f_{ij} are given by $\frac{\partial y_i^{opt}}{\partial d_j}$, such that:

$$HF = 0 \tag{1}$$

Then, $H \in \mathcal{N}(F^T)$ where $\mathcal{N}(F^T)$ is the null space of F^T. The size of F is $n_y \times n_d$ and its rank is n_d since $n_d \leq n_y$. In addition, the size of H is $n_u \times n_y$ and its rank is n_u. It can be proved that to ensure the existence of H n_y must be chosen such that $n_y \geq n_u + n_d$. In this work 2 disturbances and 5 manipulated variables were considered. Three degrees of freedom were consumed in the previous step for stabilizing the plant and one must be consumed for keeping the product quality under specification (from step 1). Therefore, the minimum n_y is 3 ($n_u = 1$ and $n_d = 2$), which is the number of outputs to be combined for the modified null space application. In other words, in this step must be chosen two loops, one for the quality control and the other for the combined measurements responsable of avoiding the disturbance effects. At this stage, the RGA is useful to detect which combination of measurements is more suitable for control with a decentralized structure (see Bristol, 1966). In this example the paring of the remaining two manipulated variables must be done, so only one element of the RGA is necessary to be calculated. In table 2 the RGA values, corresponding to the pairing $L - c$ are shown, where c is the controlled variable obtained from the lineal combination mentioned above. Then, five combinations are good candidates for implementing decentralized control structure. The chosen combination in this work was the first one for two reasons:

- the best RGA was obtained for this combination,
- all the variables that are included in this combination are easy to be measured.

The resulting control structure is shown in Fig. 1. The block H_{cont} represents the calculation of the linear combination of the variables and the PI controller that manipulates L. In Fig. 2 a flow chart is shown with all the steps included in the approach to obtain the final control structure.

4. Results

In this section the control structure developed above in steady state is tested on the dynamic model. Here, the plant under the worst disturbances scenario is used to tune the controllers guaranteeing the process specifications.
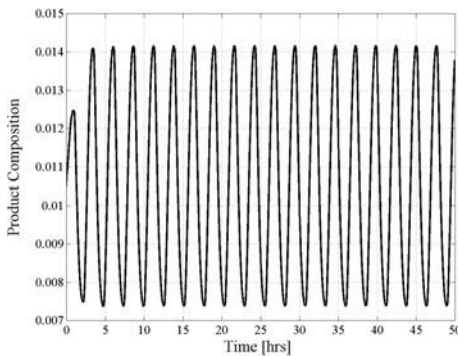
Figure 3: LATV with disturbance F_0
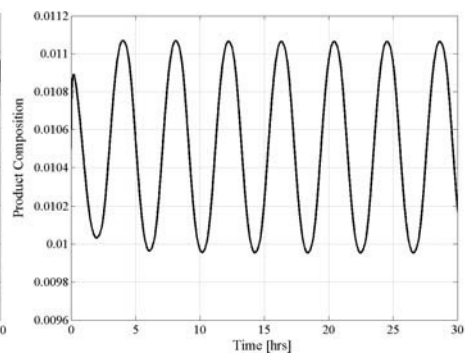
Figure 4: LATV with disturbance z_0

4.1. The Worst Case Disturbances Scenario

A worst case of dynamic disturbances was obtained by using the test described by Belanger and Luyben (1996) because of its simplicity and robustness. It consists in an Auto Tune Variation Test whose cycle tends to maximize the system's output. It is carried out connecting the process output to a device that approximates the derivative of the signal. It results in a signal that passes through zero when the process output reaches an extreme value. This signal is sent to a relay that switches the load variable when its input crosses zero magnitude. In this case the bottom product composition is one of the most important process output.

This test was done accounting the most typical disturbances: F_0 and z_0. The dynamic effects of those perturbations are presented in Fig. 3 and Fig. 4 respectively. Hence, it is possible to

evaluate which disturbances are most problematic because of their effects on the quality product composition. For the first one the frequency regimen is established at 2.46 rad/hr. It will be most critical frequency for the feed flow rate with great impact on the product quality. For the feed composition z_0 the peak of frequency is 1.54 rad/hr. Then, the control system has to be designed for rejecting these disturbances entering at those critical frequencies assumed as the worst cases.
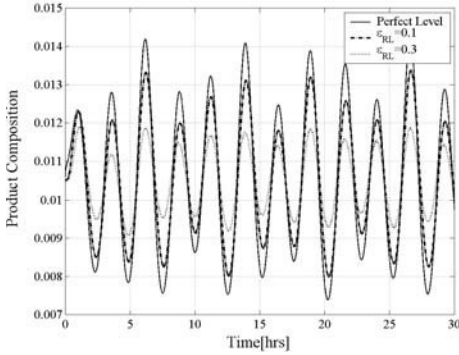
Figure 5: x_B Worst Case Response
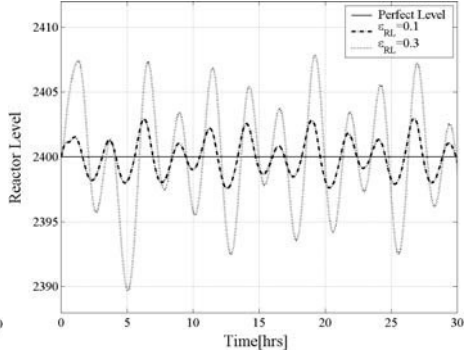
Figure 6: V_R Worst Case Response

4.2. Reactor Inventory Control Tuning

In this stage a relaxed inventory control in the reactor, as that proposed in Belanger and Luyben (1997), is adopted. The objective is to reduce the product composition variability for the disturbances analyzed here. In this work, the control tuning was done using IMC because of its simplicity on adjusting only the filter parameter (see Rivera, 2007). In the case of the reactor level the parameter is named ε_{RL}. In Fig. 5 and 6 the simulations with both disturbances affecting the plant with different values of ε_{RL} are shown . Taking into account that the composition must be $0.009 \leq x_{AB} \leq 0.012$, a part of the product will be out of specifications if $\varepsilon_{RL} = 0.1$. Otherwise, for $\varepsilon_{RL} = 0.3$ the product composition is always within the limits of the quality requirements even though it requires a larger reactor size. Hence, $\varepsilon_{RL} = 0.3$ is adopted because it gives a good trade off between controllability and product quality though higher investment on the reactor size is necessary.

In addition, step changes in the feed flow rate of about $\Delta F_0 = \pm 10\%$ and feed composition such as $\Delta z_0 = \pm 10\%$ are evaluated.
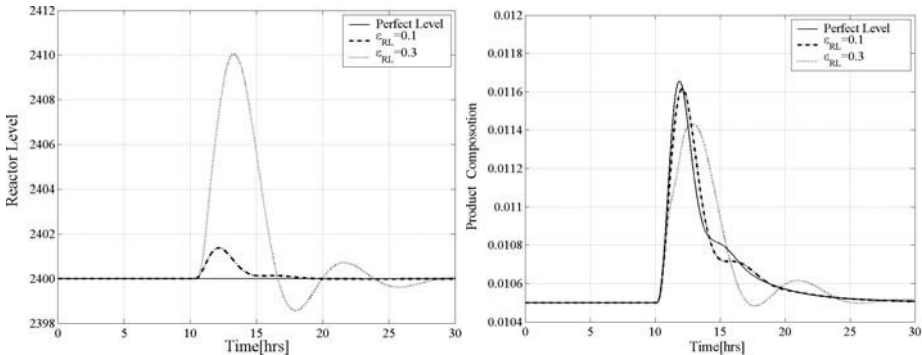
(a) Reactor Level Response

(b) Composition Product

Figure 7: Responses to Step Change in the Feed Flow Rate

In Fig. 7 the reactor level and product composition responses are shown when a step change in the feed flow rate is applied to the plant. A good regulatory behavior is noticed for the product

composition specially for the lower value of ε_{RL}. However, the opposite situation is found in the reactor level. Therefore, the same conclusion is derived from these last responses.

5. Conclusions

A new systematic methodology for synthesizing plantwide control structures has been presented. The method is based on rescuing the most useful ideas, given by the two main sources: the process-based experience and engineering judgment and a rigorous mathematical framework. This approach achieves a good decentralized structure, considering the most typical objectives for any plant, which are optimum energy management and good regulatory behavior. Even though this decision-making technique is based mainly on steady-state information, good results are achieved as it was demostrated through the simulations performed with the dynamic model. In addition, a systematic multivariable tuning method, thought to be enough robust for hadling well the most crucial disturbances, was very useful . Furthermore, this approach is helpful for deciding whether the equipment sizing is suitable for guaranteeing plant wide controllability. The satisfactory disturbance rejection was demonstrated on the integrated plant tested in the context of the worst cases. Future work will include results of the efficacy of this approach on a more complex flowsheet where a great number of possible combination for pairing variables are denoted. The solution of this large scale problem is planned to be performed with the help of data mining tools.

References

Alstad, V. and S. Skogestad (2007). Null space method for selecting optimal measurement combinations as controlled variables. *Industrial Engeenering Chemical Research* **46**, 846–853.

Belanger, P. W. and W. L. Luyben (1996). A new test for evaluation of the regulatory performance controlled variables. *Industrial Engeenering Chemical Research* **35**, 3447–3457.

Belanger, P. W. and W. L. Luyben (1997). Inventory control in processes with recycle. *Industrial Engeenering Chemical Research* **36**, 706–716.

Bristol, E. H. (1966). On a measure of interaction for multivaraible process control. *IEEE Trans. Autom. Control* **11**, 133–134.

Larsson, T. (2000). Studies On Plantwide Control. PhD thesis. Norwegain University of Science and Technology.

McAvoy, T.J. (1998). A methodology for screening level control structures in plantwide control systems. *Computers and Chemical Engineering* **22**(11), 1543–1552.

Rivera, D. E. (2007). Una metodología para la identificación integrada con el diseño de controladores imc-pid. *Revista Iberoamericana de Atutomática e Informática Industrial* **4**(4), 5–18.

Skogestad, S. (2000). Plantwide control: the search for self-optimizing control structure. *Journal of Process Control* **10**, 487–503.

Wu., K.L. and C.C. Yu (1996). Reactor / separator processes with recycles -1.candidate control structures for operability. *Comput. Chem. Eng* **20**, 1291–1316.

Control Strategy for a *Zymomonas mobilis* Bioreactor Used in Ethanol Production

Fabio C. Diehl and Jorge O. Trierweiler

Departament of Chemical Engineering - Federal University of Rio Grande do Sul, Rua Luiz Englert s/n, CEP: 900040-040 – Porto Alegre, RS, Brazil

Abstract

High oil prices, increasing focus on renewable carbohydrate-based feedstocks for fuels and chemicals have provided continuing stimulus for studies on *Zymomonas mobilis* bacteria. The ethanol production with this bacterium shows interesting nonlinear dynamic behaviors. In the continuous fermentation, there are equilibrium multiplicity, Hopf points and stability changes. Various models have been proposed in the literature (Daugulis *et al.*, 1997; Jarsebski, 1992; Ghommidh *et al.*, 1989; Jöbses *et al.*, 1986) for to describe such phenomena. In special, the Jöbses' model has captured these behaviors and revealed an attractive operating range with high substrate to product yield. This article presents a multivariable control strategy for the continuous ethanol synthesis through *Z. mobilis*, in a high yield solution branch inside the steady-state multiplicity region.
Keywords: RPN methodology, MIMO control, equilibrium multiplicity, bifurcation.

1. Introduction

Zymomonas mobilis has attracted considerable interest over the past decades as a consequence of its unique metabolism and ability to rapidly and efficiently produce ethanol from simple sugars. However, despite its apparent advantages of higher yields and faster specific rates when compared to yeasts, no commercial scale fermentations currently exist which use *Z. mobilis* for the manufacture of fuel ethanol. In the literature, the bacteria *Zymomonas mobilis* has been proposed as a more promising microorganism than conventional yeast Saccharomyces cerevisiae for industrial production of ethanol (Rogers *et al.*, 2007). Continuos glucose fermentation, with *Z. mobilis*, exhibit oscillatory behavior (Daugulis *et al.*, 1997; Jarsebski, 1992; Bruce *et al.*, 1991), and others nonlinear characteristics reported by Elnashaine *et al.* (2006), Maheca-Botero *et al.* (2006), e Garhayan e Elnashaine (2004). All these publications are based on the Jöbses's model, shortly describe in the section 2. In this paper is used the Jöbses's model to carry out the process analysis and studies. Based on this model, a suitable control strategy to ethanol fermentation by *Z. mobilis* is proposed and simulated.

2. Model Description and Analysis

Several models are reported in the literature to describe the *Z. mobilis* fermentation:

- Daugulis *et al.* (1997) – an unstructured and unsegregated model that proposes a macroscopic approximation, incorporating a dynamic specific growing rate (consider the recent process inhibition historic). This model was validated by McMellan *et al.* (1999) at low dilution rates.
- Jarsebski (1992) – consider product inhibitions and substrate limitations. It is an extension of Ghommidh *et al.* (1989) structured model, that divide the biomass in viable, unviable and dead cells.

- Jöbses *et al.* (1986) – an unsegregated-structured two-compartment representation. Divide the biomass in compartments containing specific groupings of macromolecules. It incorporates the product inhibition.

In this work was used the Jöbses' model to all analysis and projects. This model can describe several nonlinear behaviors and have a good acceptance. It is validated for low dilution rate. As a high yield operating branch appears in the steady-state multiplicity at high dilution rates, our contribution will assume that this extrapolation is acceptable. Here we are proposing a control strategy suggestion to maintain the system working at this more profitable operating region.

2.1. Jöbses's model

Since the Jöbses's model can predict a branch with higher ethanol production, which has been experimentally confirmed (at least for low dilution rates) by Elnashaie *et al.* (2006), was decided to analyze the control problem of a continuous bioreactor with the Jöbses *et al.* (1986) kinetic model, which is shortly described in this subsection.

The Jöbses's model consists of the following four differential equations:

$$\frac{dC_S}{dt} = -\left(\frac{C_S C_e}{Y_{Sx}(K_S + C_S)}\right) - m_S C_x + D(C_{So} - C_S) \tag{1}$$

$$\frac{dC_x}{dt} = \left(\frac{C_S C_e}{K_S + C_S}\right) + D(C_{xo} - C_x) \tag{2}$$

$$\frac{dC_e}{dt} = \left(k_1 - k_2 C_P + k_3 C_P{}^2\right)\left(\frac{C_S C_e}{K_S + C_S}\right) + D(C_{eo} - C_e) \tag{3}$$

$$\frac{dC_P}{dt} = \left(\frac{C_S C_e}{Y_{Px}(K_S + C_S)}\right) - m_P C_x + D(C_{Po} - C_P) \tag{4}$$

where C_S is the substrate concentration (glucose), C_x is the biomass (*Z. mobilis*), C_P is the product (ethanol), and C_e is an auxiliary variable used to lag the effect of the ethanol concentration in the kinetic model. The variables D and C_{So} represent the dilution rate and the feed substrate concentration, respectively. The C_e and the $(k_1 - k_2 C_P + k_3 C_P{}^2)$ terms make possible the model to depict the oscillatory behavior of the Hopf bifurcations. More details about the model origin can see found in Jöbses *et al.* (1986). The model parameters $(K_S, k_1, k_2, k_3, m_S, m_P, Y_{Sx}, Y_{Px})$ are shown in Jöbses *et al.* (1986).

2.2. Operational optimum

The three dimensional static/dynamic system characterization was performed using the Matcont toolbox, implemented on Matlab®, where the analyzed bifurcation parameters were D and C_{So}. The Figure 1(a) shows the results for C_P state.

The bifurcation diagram reveals triple multiplicity under some operational range. Two solutions groups are stable branches and between them, there is an unstable region. One of the stable equilibrium set present high ethanol production (region *A*) in comparison of the other stable region (region *C*). It is clear that to operate the reactor on the superior region is more profitable because the yield is bigger. The multiplicity unknowledge will lead to a trivial low yield operation. In the C_P codimension-2 analysis a Hopf bifurcation set is represented by points (●), at low dilution rates, in the lower yield field. The unstable solutions (region *B*) are linked to the superior stable branch for a saddle bifurcations line. This limit can see view in the Figure 1(b). Besides, this frontier divides the system in two partitions: the space *M*, where the multiplicity exists; and the

space *S*, where there is only a single solution. To reach the optimal profit operation, the process must be leaded in the space *M*, at the superior C_P region. As the productivity is directly dependent of *D*, the optimum operating point is close to the saddle point frontier formed by the intersection between the above stable operating range and the unstable one.
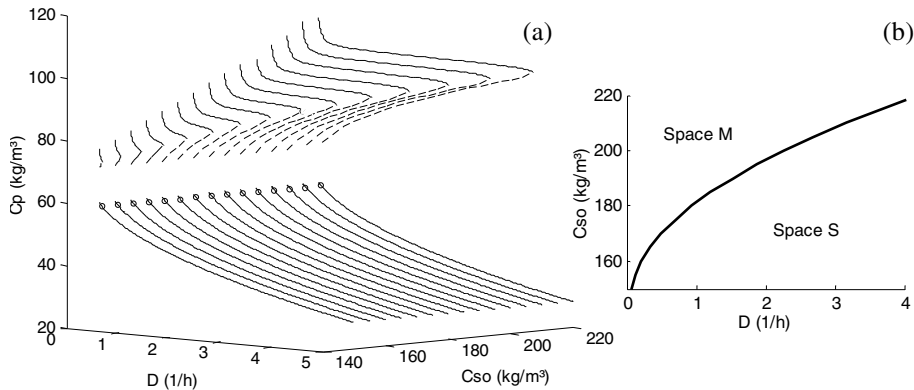
Figure 1 – Codimension-2 bifurcation (a); stability and multiplicity frontier (b).

3. Operational Controllability Analysis

Ethanol and substrate concentrations are intuitive process control variable while the natural manipulated variables are the *D* and C_{So}. It was considered the pairing $D \rightarrow C_P$ and $C_{So} \rightarrow C_S$. The operational controllability has been analyzed through the RPN methodology. The RPN (Robust Performance Number) index was introduced by Trierweiler and Engell, (1997), and the *relative* RPN (Trierweiler, 2002) indicates how potentially difficult it is for a given system to achieve the desired performance robustly. The performance with 2, 6, and 12 times faster than the open loop response and 10% overshoot have been analyzed for the high ethanol concentration branch. The RPN indicates that is possible and easy to design a robust controller with the desired performance. Additionally the nonlinearity degree has also been analyzed by the nonlinear RPN in (Trierweiler and Diehl, 2009). This analysis, reveal a predominance of static nonlinearity to region A and dynamic nonlinearity to region C. The global plant nonlinearity, region *A* and *C* together, present static and dynamic contributions. However high performance (i.e. rise time $t_R = 0.10$ *h*) becomes the global nonlinearity essentially static. Moreover, for this performance it is expected that a linear controller will be able to control the system in both stable operating regions. More details and discussions about these controllability analyses are explained in (Trierweiler and Diehl, 2009).

4. Control System

A linear PI MIMO controller was designed, based on RPN methodology tuning. The transfer matrix of the PI controller is given by

$$PI = \begin{bmatrix} -0.16\left(1 + \dfrac{1}{0.113s}\right) & 1.70\left(1 + \dfrac{1}{0.006s}\right) \\ 12.29\left(1 + \dfrac{1}{0.629s}\right) & 6.42\left(1 + \dfrac{1}{0.361s}\right) \end{bmatrix} \tag{5}$$

This quite simple controller can control all linearized models with a good performance as it is shown in Figure 2. Ten linearized models (five belonging to region *A* and other five belonging to region *C*) are controlled by the PI MIMO, proving that it is possible to control the global plant with a linear controller. The control system is tested in a nonlinear Jöbses's model for different set points at region *A*. The results are showed in Figure 3. The performance is very good, but in a real process application the C_{So} difficultly would be prepared instantly. For this reason a first order transfer function was inserted before the C_{So} bias sum (representing a mixer for example). This function adds 10 minutes time delay distributed in a first order curve. The Figure 3 shows the control system robustness characteristic, because no new design was performed.
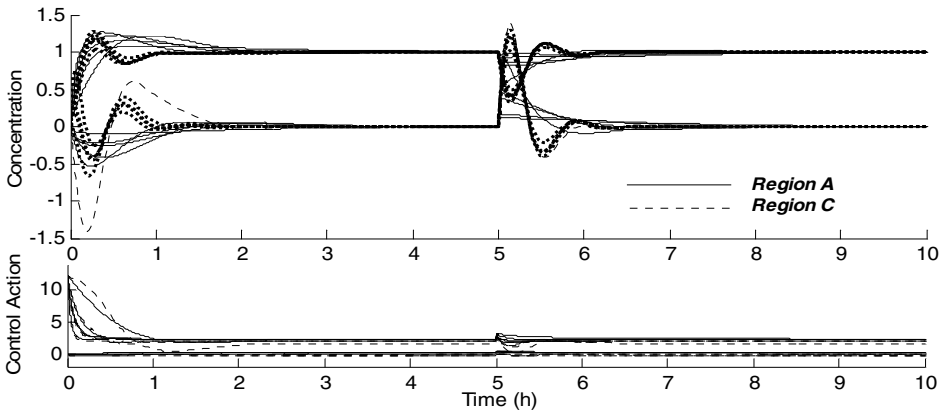
Figure 2 – PI MIMO: global plant control.
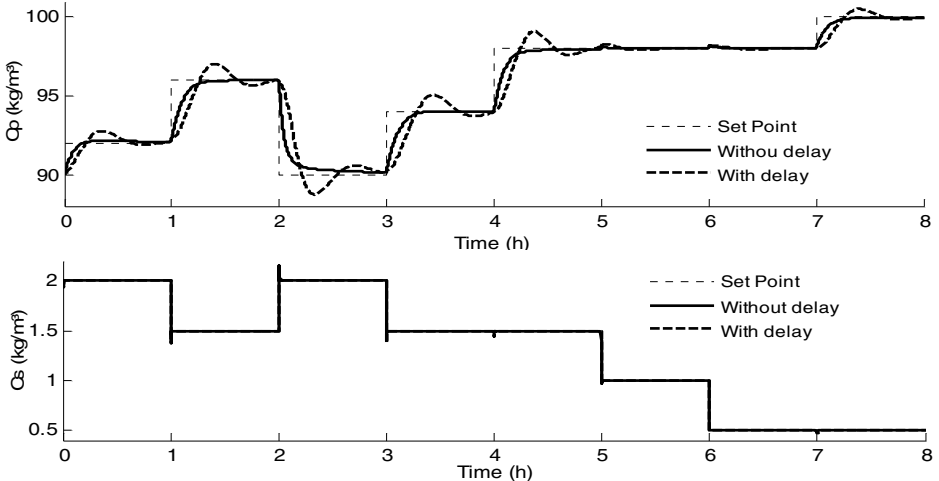
Figure 3 – Multivariable linear controller in the nonlinear model.

According to previously discussed, the operational optimum is the stability frontier. It is desirable to maintain the process at the space *M*, in the high yield field (region *A*). However immeasurable disturbance and/or unfeasible set points can lead the reactor convergence to the region *C*. To avoid this situation the control system actuation requires an operational limitation. The Figure 4 illustrates the restriction strategy. First the *D* and C_{So} axes are transposed to a new origin point at the stability frontier. Through a rotation matrix *R* the actions are recalculated to a rotated coordinates. The space *M* actuation is guaranteed since the C_{so} variable is positive. In the rotation matrix *R* the β angle is the angular coefficient of the tangent line to a stability limit point. The successful strategy results can be viewed in the Figure 5, for an unfeasible C_P set point.

$$R = \begin{bmatrix} \cos(\beta) & -\sin(\beta) \\ \sin(\beta) & \cos(\beta) \end{bmatrix}$$
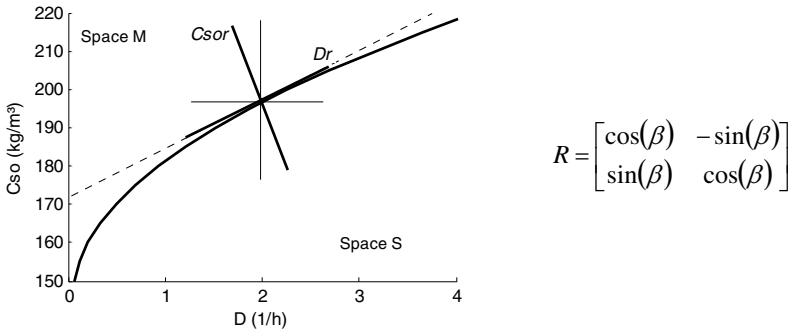
Figure 4 – Limit actuation problem solution.
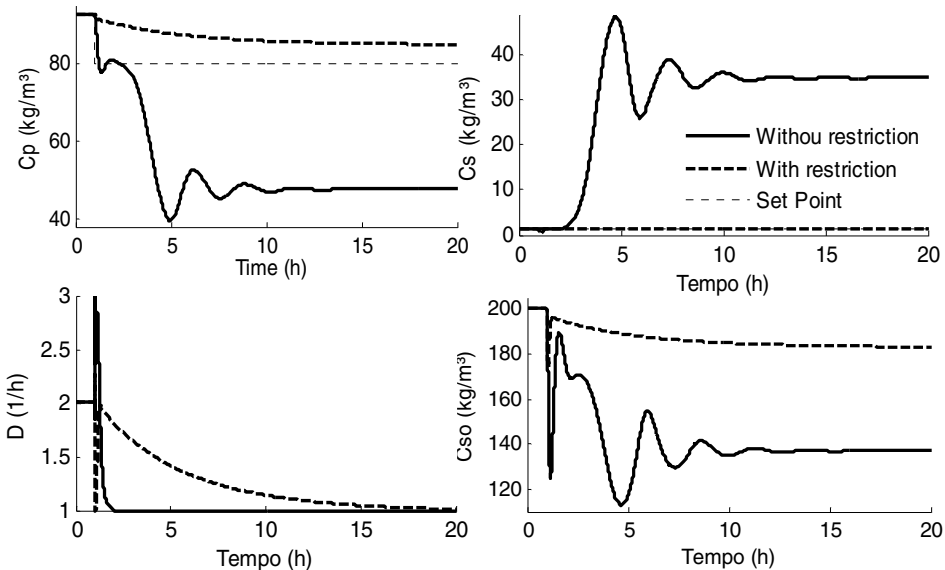
Figure 5 – Control variables without and with restrictions.

5. Conclusions

This contribution proposes a control strategy to high ethanol yield in an equilibrium multiplicity region. All analyses were based in the Jöbses's model. However, few experimental results have been confirmed the predicted results at high dilution rates so that more studies should be performed to the model validation. The codimension-2

bifurcation analysis reveals the multiplicity extension behaviour at Jöbses's model. A high stable yield operating region appears on the triple operational regions. This branch is linked with an unstable branch and the stability limit curve formed represents the operational optimum. So a control system was designed to maintain the process on this optimum.

A MIMO control structure was proposed, with the pairing $D \rightarrow C_P$ and $C_{So} \rightarrow C_S$. Based on the RPN methodology the robust performance is easily attainable. The fast controller performance minimizes the dynamic nonlinearity, becoming the global nonlinearity essentially static. The nonlinear analysis revels in a possibility to control the global process with a linear controller.

A multivariable PI controller was designed based on the RPN methodology. This controller could control the system in both stable operational regions (global control). Yet, it was capable to hold the operation on the region A, when the nonlinear model is simulated.

It was presented a safe strategy to maintain the operation at high yield region A. This strategy simply restricts the variable control actuation area. For this a rotation matrix change the action coordinates for satisfy a limit criterion. When Cso_R is bigger than zero the system is on the multiplicity space M.

References

A. Maheca-Botero, P. Garhyan, S.S.E.H. Elnashaine (2006) Non-linear characteristics of a fermentor for ethanol production and their implications. *Nonlinear Analysis: Real World Applications* 7, 432-457.

A.B. Jarzebski (1992) Modelling of oscillatory behaviour in continuous ethanol fermentation. *Biotechnology Letters* 2, 137-142.

A.J. Daugulis, P.J. Mclellan, J. Li (1997) Experimental investigation and modeling of oscillatory behavior in the continuous culture of *Zymomonas mobilis*. *Biotechnology and Bioengineering,* 56**,** 99-105.

C. Ghommidh, J. Vaija, S. Bolarinwa, J.M. Navarro (1989) Oscillatory behaviour of Zymomonas in contínuos cultures: a simple stochastic model. *Biotechnology Letters* **9**, 659-664.

I.M.L. Jöbses, G.T.C. Egberts, K.C.A.M. Luyben, J.A. Roels, (1986) Fermentation kinetics of *Zymomonas mobilis* at high ethanol concentrations: Oscillations in continuous cultures. *Biotechnology and Bioengineering,* 28**,** 868-877.

J.O. Trierweiler (2002) Application of the RPN Methodology for Quantification of the Operability of the Quadruple-Tank Process. *Brazilian Journal of Chemical Engineering,* 19**,** 195-206.

J.O. Trierweiler, F.C. Diehl (2009) Analysis, control, and operational optimization of a *Zymomonas mobilis* reactor with equilibrium multiplicity. *Submitted to ADCHEM 2009*

J.O. Trierweiler, S. Engell (1997) The Robust Performance Number: A New Tool for Control Structure Design. *Computer Chemical Engineering* 21**,** 237-243.

L.J. Bruce, D.B. Axford, B. Ciszek, A.J. Daugulis (1991) Extractive fermentation by *Zymomonas mobilis* and the control of oscillatory behavior. *Biotechnology Letters* 13, 291-296.

P. Garhyan, S.S.E.H. Elnashaie (2004) Utilization of mathematical models to investigate the bifurcation and chaotic behavior of ethanol fermentors. *Mathematical and Computer Modelling* 39, 381-427.

P. Rogers, Y. Jeon, K. Lee, H. Lawford (2007) *Zymomonas mobilis* for Fuel Ethanol and Higher Value Products. *Biofuels.* Springer Berlin / Heidelberg.

P.J. McLellan, A.J. Daugulis, J. Li (1999) The incidence of oscillatory behavior in the continuous fermentation of *Zymomonas mobilis*. *Biotechnology Progress* 15, 667-680.

S.S.E.H. Elnashaie, Z. Chen, P. Garhyan, P. Prasad, A. Mahecha-Botero (2006) Practical Implications of Bifurcation and Chaos in Chemical and Biological Reaction Engineering. *International Journal of Chemical Reactor Engineering 4*.

Robust Output-Feedback Nonlinear Model Predictive Control Using High-Gain Observers

Rui Huang[a], Sachin C. Patwardhan[b], Lorenz T. Biegler[a]

[a] *Department of Chemical Engineering, CMU, Pittsburgh, PA, USA 15213.*
[b] *Department of Chemical Engineering, IIT Bombay, Mumbai, India, 400076.*

Abstract

A robust output-feedback nonlinear model predictive control (NMPC) system is proposed by combining state-feedback NMPC with a certain type of high-gain observer. We show that robust stability of this formulation deteriorates from its state-feedback NMPC counterpart because of observation error. Then we demonstrate this idea through simulation studies of a CSTR example and offset free regulatory behavior is obtained because the estimated error is used to correct model-mismatch in the controller.
Keywords: Robust controller, Output-feedback system, High-gain observer.

1. Introduction

Nonlinear model predictive control (NMPC) has received considerable attention over the past years. Schemes to guarantee the stability of the closed-loop systems with state-feedback NMPC have been widely studied (Limon et al. 2008 & Mayne et al, 2000). However, the assumption that the plant is available for measurement usually does not hold in practice. Hence, the output-feedback NMPC controller is required by integrating a state-feedback NMPC controller with a state observer. For linear systems, the well-known separation principle guarantees that the closed-loop system is nominally stable if both the controller and the state observer are stable. For nonlinear systems, the issue of stability of the closed-loop systems is still open due to the lack of general valid separation principles. In order to analyze the nominal stability for certain nonlinear output-feedback systems, two possibilities have been explored: 1) developing a certain equivalent separation principle and designing observers accordingly (Atassi & Khalil, 1999, Findeisen. et al, 2003); 2) considering the observer error in the NMPC controller (Michalska & Mayne, 1995. Magni et al, 2004).

The above formulations focus on the nominal closed-loop stability or stability in the presence of vanishing perturbations. In this work, we propose a robust output-feedback NMPC based on a high-gain observer. We show the robust stability analysis for the proposed formulation with bounded observation errors. Section 2 starts with the introduction of the output-feedback system and stability analysis. Section 3 presents simulation examples of continuous stirred tank reactor (CSTR). We also demonstrate that the proposed formulation generates offset-free regulatory and servo behavior under moderate perturbations in model parameters. Section 4 concludes the paper.

2. Problem Formulation

2.1. State-feedback NMPC and stability

In this work, we consider the dynamic model of a plant with output.

$$x_{k+1} = f(x_k, u_k, \theta_k) \qquad y_k = Cx_k, \quad k \geq 0 \tag{1}$$

where $x_k \in \mathfrak{R}^{n_x}$, $u_k \in \mathfrak{R}^{n_u}$, $y_k \in \mathfrak{R}^{n_y}$ and $\theta_k \in \mathfrak{R}^{n_\theta}$ are the plant states, controls, outputs and uncertainty parameters respectively, defined at time steps t_k with integers $k \geq 0$. Without losing generality, we assume that the given plant (1) has an equilibrium point at the origin, that is $f(0,0,0) = 0$.

Given $x(k)$, the current state value at time step t_k, the state-feedback NMPC formulation can be described in the following discretized form:

$$V(x_k) := \min \ \textstyle\sum_{j=0}^{N-1} l(z_{k+j}, u_{k+j}) + F(z_{k+N}) \tag{2a}$$

$$\text{s.t.} \qquad z_{k+j+1} = f(z_{k+j}, u_{k+j}, 0), \quad j = 0, \ldots, N-1 \tag{2b}$$

$$z_k = x_k, z_{k+j} \in \mathcal{X}, z_{k+N} \in \mathcal{X}_f, u_{k+j} \in \mathcal{U} \tag{2c}$$

where N is the finite time horizon. z_{k+j} is the sequence of the predicted state variables, and u_k is the calculated control action based on the plant state x_k at time step t_k. Note the uncertainty parameter is 0 in the controller, introducing the plant-model mismatch. The calculated state-feedback control law from (2) can be written as $u_k = h(x_k)$, and the plant state at the next time step t_{k+1} can be expressed as $f(x_k, h(x_k), \theta)$.

The closed-loop stability with the state-feedback control law has been widely studied. Here we summarize the key definitions and analysis developed in Limon et al 2008. Here, $|\cdot|$ denotes the Euclidean vector norm in $\mathfrak{R}^n$ or the associated matrix norm. For a given sequence w_k, we denote $\|w\| \triangleq \sup_{k \geq 0}\{|w_k|\}$. For two $\mathcal{K}$ functions τ_1 and τ_2, we define $\tau_1 \circ \tau_2(s) \triangleq \tau_1(\tau_2(s))$.

Lemma 1: *If $f(x,y)$ is a uniformly continuous function in both $x \in A$ and $y \in B$, then there exist $\mathcal{K}$ functions τ_1 and τ_2, such that $|f(x_1,y_1) - f(x_2,y_2)| \leq \tau_1(|x_1 - x_2|) + \tau_2(|y_1 - y_2|)$.*

Theorem 1: *(Limon et al 2008) Assume function $f(x, h(x), \theta)$ is uniformly continuous in θ, and the plant (1) is nominally asymptotically stable. If there exists a uniformly continuous Lyapunov function $V(x)$ for plant (1), then the closed-loop system with state-feedback controller is Input-to-State stable (ISS).*

2.2. Output-feedback NMPC and stability

Consider now a state observer for plant (1),

$$\hat{x}_{k+1} = g(\hat{x}_k, y_k, h(\hat{x}_k)), \quad k \geq 0 \tag{3}$$

where $\hat{x}_k$ is the state estimated from the outputs at time step t_k. Hence, the initial condition in the NMPC formulation is modified as $z_k = \hat{x}_k$, instead of $z_k = x_k$ in (2c). The calculated output-feedback control law is $u_k = h(\hat{x}_k)$, and the plant state at t_{k+1} can be modified as $f(x_k, h(\hat{x}_k), \theta)$.

In order to establish the robust stability of the output-feedback NMPC formulation, we make use the following assumptions.

Assumption 1 (Observer assumption)

   a) *the initial observer error is bounded by a positive constant e, i.e. $|\hat{x} - x| \leq e$;*

   b) *the future observer error satisfies $|\hat{x}_k - x_k| \leq \rho|e|\eta^{-k} + \beta(\|\theta\|)$, where $\rho \geq 0, \eta > 1$ are constants, β is a $\mathcal{K}$ function.*

   c) *the observer is a high-gain observer.*

Lemma 2: *There exists a constant $\epsilon \geq 0$ such that $\rho|e|\eta^{-k} \leq \epsilon$. As a result $|\hat{x}_k - x_k| \leq \epsilon + \beta(\|\theta\|)$.*

Lemma 3: *If $|\hat{x} - x| \le e$, then $|x| - e \le |\hat{x}| \le |x| + e$. For a $\mathcal{K}$ function $\alpha(\cdot)$, there are other $\mathcal{K}$ functions $\alpha_L(\cdot)$, $\alpha_U(\cdot)$ and positive constant c_1, c_2, such that $\alpha_L(|x|) - c_1 \le \alpha(|x| - e) \le \alpha(|\hat{x}|) \le \alpha(|x| + e) \le \alpha_U(|x|) + c_2$. Similarly if $|\hat{x} - x| \le \epsilon + \beta(\|\theta\|)$, we can find other $\mathcal{K}$ functions $\hat{\alpha}_L(\cdot)$, $\hat{\alpha}_U(\cdot)$, $\beta_L(\cdot)$, $\beta_U(\cdot)$ and positive constants c_3, c_4, such that $\hat{\alpha}_L(|x|) - \beta_L(\|\theta\|) - c_3 \le \alpha(|\hat{x}|) \le \hat{\alpha}_U(|x|) + \beta_U(\|\theta\|) + c_4$.*

Robust stability of this output-feedback NMPC can be established by the following:

Theorem 2: *Assume function $f(x, h(\hat{x}), \theta)$ is uniformly continuous in θ and x, and system (1) is nominally asymptotically stable and the observer satisfies Assumption 1. If there exists a uniformly continuous Lyapunov function $V(x)$, then the closed-loop system with output-feedback controller is Input-to-state practical stable (ISpS).*

Proof: The analysis is similar as the proof for Theorem 2 in Limon et al 2008, but we need to consider the observer error. From the continuity of $V(x)$ and $f(x, h(\hat{x}), \theta)$, there exist $\mathcal{K}$ functions α, σ_V, σ_f and σ_θ such that $V(f(\hat{x}, h(\hat{x}), 0)) - V(\hat{x}) \le -\alpha(|\hat{x}|)$, $V(\hat{x}) - V(x) \le \sigma_V(|\hat{x} - x|)$, and $V(f(x, h(\hat{x}), \theta)) - V(f(\hat{x}, h(\hat{x}), 0)) \le \sigma_V(|f(x, h(\hat{x}), \theta) - f(\hat{x}, h(\hat{x}), 0)|) \le \sigma_V \circ (\sigma_f(|\hat{x} - x|) + \sigma_\theta(|\theta|))$. As a result:

$$V(f(x, h(\hat{x}), \theta)) - V(x) = V(f(\hat{x}, h(\hat{x}), 0)) - V(\hat{x}) + V(\hat{x}) - V(x)$$
$$+ V(f(x, h(\hat{x}), \theta)) - V(f(\hat{x}, h(\hat{x}), 0))$$
$$\le -\alpha(|\hat{x}|) + \sigma_V(|\hat{x} - x|) + \sigma_V \circ (\sigma_f(|\hat{x} - x|) + \sigma_\theta(|\theta|)) \qquad (4)$$

Now, we need to ensure that the closed-loop system is robustly stable for both the initial stage of the observer when $|\hat{x} - x| \le e$ and later stages when the observer error is corrupted by the uncertainties, i.e. $|\hat{x} - x| \le \epsilon + \beta(\|\theta\|)$.

If $|\hat{x} - x| \le e$, from Lemma 3, we have $-\alpha(|\hat{x}|) \le -\alpha_L(|x|) + c_1$. In addition, we can find a constant $c_5 \ge 0$ such that $\sigma_V(e) + \sigma_V \circ \sigma_f(e) + c_1 \le c_5$ and a $\mathcal{K}$ function σ_2 such that $\sigma_V \circ \sigma_\theta(|\theta|)) \le \sigma_2(|\theta|)$. Consequently equation (4) leads to:

$$V(f(x, h(\hat{x}), \theta)) - V(x) \le -\alpha(|\hat{x}|) + \sigma_V(|\hat{x} - x|) + \sigma_V \circ (\sigma_f(|\hat{x} - x|) + \sigma_\theta(|\theta|))$$

$$\le -\alpha_L(|x|) + \sigma_2(|\theta|) + c_5 \qquad (5)$$

Then the closed-loop system is ISpS stable for the initial observer error.

If $|\hat{x} - x| \le \epsilon + \beta(\|\theta\|)$, we have $-\alpha(|\hat{x}|) \le -\hat{\alpha}_L(|x|) + \beta_L(\|\theta\|) + c_3$, then we can pick a constant $c_6 \ge 0$ and a $\mathcal{K}$ function σ_3 such that $\beta_L(\|\theta\|) + c_3 + \sigma_V(|\hat{x} - x|) + \sigma_V \circ \sigma_f(|\hat{x} - x|) + \sigma_V \circ \sigma_\theta(|\theta|) \le \sigma_3(\|\theta\|) + c_6$. Then equation (4) leads to:

$$V(f(x, h(\hat{x}), \theta)) - V(x) \le -\alpha(|\hat{x}|) + \sigma_V(|\hat{x} - x|) + \sigma_V \circ (\sigma_f(|\hat{x} - x|) + \sigma_\theta(|\theta|))$$

$$\le -\hat{\alpha}_L(|x|) + \beta_L(\|\theta\|) + c_3 + \sigma_V(|\hat{x} - x|) + \sigma_V \circ \sigma_f(|\hat{x} - x|) + \sigma_V \circ \sigma_\theta(|\theta|)$$

$$\le -\hat{\alpha}_L(|x|) + \sigma_3(\|\theta\|) + c_6 \qquad (6)$$

Then the closed-loop system is ISpS stable for the later stage when observer error is corrupted by the uncertainties. ?

Remark: From the analysis, we see that the ISS stability of state-feedback system deteriorates to ISpS stability of output-feedback system, because of the corruption of observer error. For the output-feedback system, all the past uncertainty signals affect the closed-loop stability, while only the current uncertainty signal plays a role in the

robust stability for the state-feedback system. In addition, stronger assumption is required to guarantee the robust stability of the output-feedback system.

2.3. EKF as the High-gain observer stability

In this section we consider an extended Kalman filter (EKF) developed by Reif & Unbehauen (1999) as the state observer, satisfying Assumption 1.

$$x_{k+1}^- = f(\hat{x}_k, u_k, 0) \quad \hat{x}_k = x^- + K_k[y_k - Cx_k^-] \tag{7}$$

where x_k^- and $\hat{x}_k$ are called the prior and posterior estimate. K_k is the Kalman gain calculated from:

$$K_k = P_k^- C^T [C P_k^- C^j + R]^{-1} \quad P_{k+1}^- = \alpha^2 A_k P_k^+ A_k^T + Q \tag{8a}$$

$$P_k^+ = [I - K_k C] P_k^- \tag{8b}$$

where $A_k = \dfrac{\partial f}{\partial x}(\hat{x}_k, u_k, 0)$, $G_k = \dfrac{\partial f}{\partial \theta}(\hat{x}_k, u_k, 0)$ are the linearizations of the model. Q and R are symmetric positive definite matrices. $\alpha \geq 1$ is a exponential data weight to control the convergence rate of the EKF. Hence this EKF formulation is a high-gain observer. If $\alpha = 1$, this formulation reduces to the conventional EKF. Let $\varphi(\cdot, \cdots)$ be the higher order term, the residual of the EKF can be defined by the following equations:

$$f(x_k, u_k, \theta) - f(\hat{x}_k, u_k, 0) = A_k[x_k - \hat{x}_k] + G_k \theta_k + \varphi(x_k, \hat{x}_k, u_k, \theta_k) \tag{9}$$

Defining estimation error as $\zeta_k = x_k - \hat{x}_k$, we have from (1), (7), (8) and (9),

$$\zeta_{k+1} = (I - K_{k+1} C)[A_k \zeta_k + r_k] \tag{10}$$

where $r_k = \varphi(x_k, \hat{x}_k, u_k, \theta_k) + G_k \theta_k$

Theorem 2: *Consider the EKF as stated by equations (7) and (8) and let the following assumptions hold:*

   a)   *There are positive numbers $\bar{a}, \underline{p}, \bar{g}$ and $\bar{p}$ such that $|A_k| \leq \bar{a}$, $|G_k| \leq \bar{g}$, $\underline{p}I \leq P_k^- \leq \bar{p}I$ and $\underline{p}I \leq P_k^+ \leq \bar{p}I$, $\forall k \geq 0$.*
   b)   *A_k is nonsingular for $k \geq 0$.*
   c)   *$\theta_k \in \Omega_\theta \subset \Re^{n_\theta}$ where Ω_θ is a compact set.*
   d)   *There are positive real numbers $\kappa, \gamma_0, \mathcal{K}$ functions $\delta(\|\theta\|)$ and $\mu(\|\theta\|)$, such that $|\varphi(x_k, \hat{x}_k, u_k, \theta_k)| \leq \kappa |x - \hat{x}_k|^2 + \delta(\|\theta\|)$ whenever $|x - \hat{x}_k| \leq \mu(\|\theta\|) + \gamma_0 \leq \bar{\varepsilon}$.*

Then with Assumption 1.a, there exist $\rho \geq 0, \eta > 1$ and $c_e \geq 0$ are constants, β is a $\mathcal{K}$ function, such that the error sequence ζ_k defined in equation (10) behaves according to $|\zeta_k| \leq \rho |\zeta_0| \eta^{-k} + \beta(\|\theta\|)$.

The proof follows the similar line from the proof in Reif & Unbehauen (1999). With this theorem, we can see Assumption 1.b and 1.c are true for this EKF.

2.4. Output-feedback NMPC with EKF

To achieve offset free closed-loop behavior, we propose to carry out multi-step predictions by explicitly using observer errors for future predictions as follows:

$$z_{k+j+1} = f(z_{k+j}, u_{k+j}, 0) + K_k \beta_{k+j}; \quad y_{k+j} = C z_{k+j} + \eta_{k+j}, j = 0, \ldots, N-1 \tag{11a}$$

$$\beta_{k+j+1} = \beta_{k+j}, \eta_{k+j+1} = \eta_{k+j}, z_k = \hat{x}_k, \ \beta_k = y_k - C x_k^-, \ \eta_k = y_k - C \hat{x}_k \tag{11b}$$

The predictive controller is then formulated as

$$\min_u \sum_{j=1}^{N-1}[E_{k+j}^T W_E E_{k+j}] + \sum_{i=0}^{q-1} \Delta u_{k+i}^T W_{\Delta u} u_{k+i} + E_{k+N}^T W_\infty E_{k+N}$$

$$s.t. \; z_{k+j} \in \mathcal{X}, z_{k+N} \in \mathcal{X}_f, u_{k+i} \in \mathcal{U}, j = 0, \ldots, N-1, \; i = 0, \ldots, q-1, \text{(11a,b)}$$

$$u_{k+j} = u_{k+t}, \text{ for } t_i \le j < t_{i+1}, \; t_0 = 0 < t_1 < t_2 < \ldots < t_{q-1} = N-1. \text{ (12)}$$

where $E_{k+j} \triangleq (y_{k+j} - y_r)$, $W_E, W_{\Delta u}$ and W_∞ are positive semi-definite matrices.

3. Simulation Studies

We consider a simulated NMPC scenario with a nonlinear CSTR model represented by the following differential equations:

$$\frac{dz_c}{dt} = (z_c - 1)/u_1 + k_0 z_c \exp(-E_a/z_T) \tag{13a}$$

$$\frac{dz_T}{dt} = (z_T - z_T^f)/u_1 + k_0 z_c \exp(-E_a/z_T) + \nu u_2(z_T - z_T^{cw}) \tag{13b}$$

This system involves two states $z = [z_c, z_T]$ corresponding to dimensionless concentration and temperature, and two manipulated inputs, corresponding to the inverse of dilution rate (u_1) and cooling water flow rate (u_2). The model parameters are $z_T^{cw} = 0.38$, $z_T^f = 0.395$, $E_a = 5$, $\nu = 1.95 \times 10^1$ and k_0 is an uncertainty parameter in the plant with nominal value $k_0 = 300$ in the model. The system is operated at a stable steady state $z_c = 0.1247$ and $z_T = 0.74070$ corresponding to $u_1 = 20$ and $u_2 = 378$. The NMPC is formulated using $W_E = W_\infty = \text{diag}[1 \times 10^6 \; 1 \times 10^5]$, prediction horizon $N = 20$, control horizon $q = 5$, with input blocking and each block equals to 4 samples, and sampling time is 1. W_{2u} is chosen as a null matrix. The EKF is tuned with $Q = \left[\frac{\partial f}{\partial u}|_{z_{ss}, u_{ss}}\right] \bar{Q} \left[\frac{\partial f}{\partial u}|_{z_{ss}, u_{ss}}\right]^T$, where $\bar{Q}$ is chosen to be the possible variations in the inputs $\text{diag}[6.25, 0.04]$, and $R = \text{diag}[1 \times 10^{-6}, 1 \times 10^{-6}]$ which is the possible covariance of the outputs. To ensure that EKF works as a high gain observer, α in equation (8a) is chosen equal to 2.5.

Figure 1 presents the variation of controlled output and state estimation errors generated in response to a sequence of +/- 40 % step changes in model parameter k_0 (see Figure 2) for two different scenarios, 1) without output noises, 2) outputs are corrupted with white noise with standard deviation of $[1 \times 10^{-3}, 1 \times 10^{-3}]^T$. The corresponding variation of manipulated inputs is presented in Figure 2. It is clear from Figure 1 that, each time after a step change is introduced in k_0, the state estimator errors generated by the high gain observer settle to a constant value in a few samples. Moreover, the high gain observer's performance is hardly influenced by the tuning of the controller. It may be noted that the proposed output-feedback NMPC formulation eliminates the offset in both the cases.

4. Concluding Remarks

The purpose of this paper is to point out that although there are no general separation principles for nonlinear systems with plant-model mismatches, the closed-loop stability can still be ensured by using certain types of observers, and the controller performance can be improved by considering the observer error and tuning the observer. In future, this technique will be implemented on large scale applications and other types of observers will also be studied.
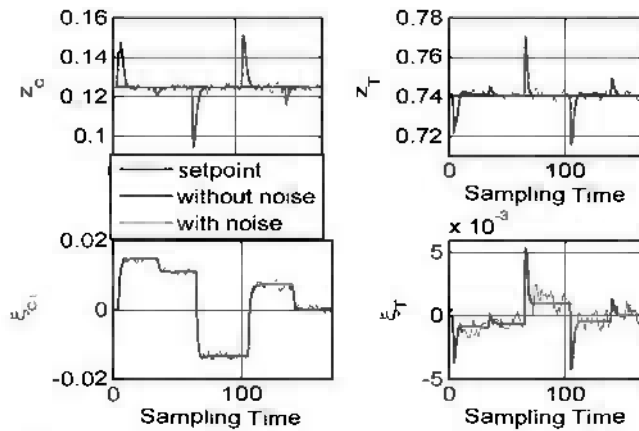
Figure 1: Variation of controlled outputs and state estimation errors
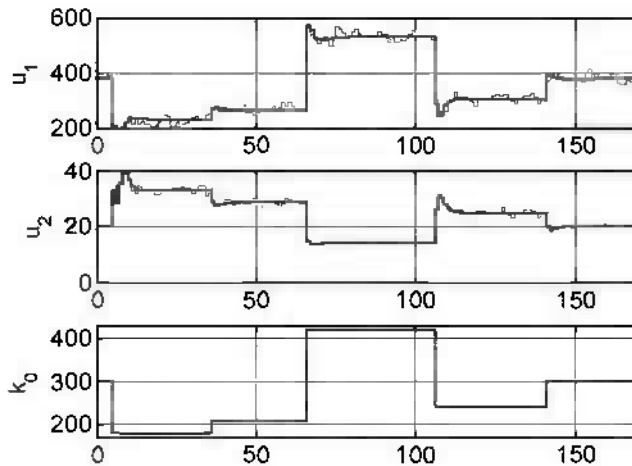
Figure 2: Variation of manipulated input and parameter disturbance

References

A Atassi & H. Khalil, 1999. "A separation principle for the stabilization of a class of nonlinear systems", IEEE Trans. on Auto. Contr. 44: 1672.

R. Findeisen, L. Imsland, F. Allgöwer & B.A. Foss, 2003. "State and output feedback nonlinear model predictive control: an overview". European J. of Control. 9: 2.

D. Limon. T. Alamo, D. Raimondo, D.M. Peña, J. Bravo & E. Camacho, 2008. "Input-to-state stability: a unifying framework for robust model predictive control". In Assessment and Future Directions of Nonlinear Model Predictive Control. Springer, in press.

L. Magni, G.De Nicolao & R. Scattolini, 2004. "On the stabilization of nonlinear discrete-time systems with output feedback". International J. of Robust and Nonlinear Control. 14:1379.

D Mayne. J. Rawlings, C. Rao & P.O.M. Scokaert, 2000. "Constrained model predictive control: stability and optimality". Automatica, 36: 789.

H. Michalska & D. Mayne, 1995. "Moving horizon observers and observer-based control", IEEE Trans. on Auto. Contr, 40: 995.

K. Reif & R. Unbehauen, 1999. "The extended Kalman filter as an exponential observer for nonlinear systems", IEEE Trans. on Auto. Contr, 47: 2324.

1617

Scilab/Scicos: An Alternative Tool for Real-Time Monitoring and Advanced Control of Fieldbus Industrial Systems

Thiago V. Costa, Ana M. F. Fileti, Flávio V. Silva

LCAP/DESQ - Department of Chemical Systems Engineering, School of Chemical Engineering, University of Campinas, Av. Albert Einstein, 500, CP 6066, CEP13083-970, Campinas-SP, Brazil

Abstract

The present work deals with real-time data acquisition and advanced control evaluation utilizing the open source scientific platform Scilab. Implementation and visualization of online data with the Scicos toolbox and utilization of OPC technology were discussed. The feasibility and effectiveness of the proposed methodology was shown by means of application of fuzzy controllers to a bromelain enzyme precipitation process instrumented with Foundation Fieldbus devices. Results confirmed Scilab/Scicos suitable for HMI and control systems applied in small industrial applications.

Keywords: real-time control, OPC connection, precipitation process, fuzzy logic.

1. Introduction

Personal computers (PC's) have become a reliable option to process control, as they are practical, have low cost and high computational performance. While dedicate machines have advantages over personal computers like optimized operation and specific hardware system for control, PC's have a greater range of programming languages (*C, Java* etc.) and development tools (Seborg et al, 2004). The greater availability of computer tools can improve the diversification of current industrial control loops while the real-time implementation and evaluation of the control algorithms in a PC control application can be performed with minimal effort through an OPC communication. Based on Microsoft *OLE/COM* technology, OPC (ole for process control) consists of standardized protocols to easy information flow and interoperability among automation, control loops and software systems. These common specifications eliminate the need of proprietary hardware vendor drivers, allowing the development of control systems, human machine interfaces (*HMI*) and supervisory control and data acquisition (*SCADA*) systems in scientific softwares, such as Scilab, making use of its mathematical toolboxes in the formulation of real-time advanced control algorithms, as proposed by Peng and Ma (2006).

Scilab is a free open source scientific platform for numerical computations developed in 90's at *INRIA* (*"Institut National de Recherche en Informatique et Automatique"*), maintained since 2003 by a consortium composed by companies and academic institutions (www.scilab.org). Scilab is not properly intended for real-time control; it's a simulation and analysis program. Even though, it is fast enough to be considered for soft real-time applications, which means that a suitable control loop could be evaluated in not so fast sampling times (more than 1s for example). This fortunately covers most of industrial chemical processes. For the implementation and

visualization of online data this work proposes the utilization of the toolbox Scicos, a block diagram editor primarily intended for dynamic, discrete and hybrid systems simulation (Najafi et al, 2003; Nikoukhah and Steer, 2006). Drag and drop and interconnection of user defined and pre-programmed function blocks allow algorithm implementation in an easier and straightforward manner. Interface and interoperability between the process data and the user is done in the same fashion with display boxes and scale widgets resulting in clear and functional supervisory screens.

To show experimentally the benefits of this methodology, a bromelain precipitation plant was real time monitored and controlled by Scilab/Scicos. Temperature control of the precipitation tank was performed by advanced controllers based on PI, PI+D and PI+PD forms of feedback digital fuzzy logic controllers using Scilab's mathematical routines. Online closed loop tests evaluating the fuzzy logic controllers in comparison with classical PID controller were developed and accompanied by performance indexes like *ITAE* (integral time average absolute error criteria), energy consumption and system behavior aspects.

2. Scilab/Scicos OPC Communication

The Scilab toolbox for OPC connection was developed by Peng and Ma (2006). The functions were compiled in a dynamic linking library (*DLL*) and linked with Scilab, using its interface to *C* and *FORTRAN* codes. Table 2.1 shows the toolbox functions available from the *OPC.dll* file.

Table 2.1 OPC toolbox functions (adapted from Peng and Ma, 2006)

opc_connect	connect OPC server
opc_disconnect	disconnect OPC server
opc_add_group	add a group to OPC server
opc_add_item	add OPC items to OPC group
opc_item_write	write OPC items
opc_item_read	read OPC items

The OPC elements arrangement is shown in Figure 2.1. After the connection with the OPC server, a *Group* responsible for *Items* organization (tags or memory addresses) is added. *Items* are placed in a string vector and called by their position.
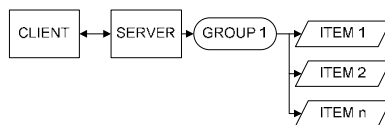
Figure 2.1 *OPC* client/server connection.

Scicos interface with field instruments is carried out in a quite easy procedure. First, the diagram has to be configured for *real-time scaling* in *Simulation/Setup* option. Then, the instructions for *connection, group* and *items* adding are placed in *Diagram/Context*. These instructions will run every time the diagram is loaded. Reading and writing of process data is performed by Scilab code blocks *(scifuncs)* which are placed in the diagram loop.

3. Bromelain Precipitation Process

The bromelain precipitation process (see Figure 3.1a) consists in a jacketed fed-batch stirred tank, instrumented with Fieldbus Foundation intelligent devices, where bromelain enzyme is recovered from pineapple pulp by precipitation with alcohol at low temperatures (5°C).

The tank (1000ml of nominal capacity) is initially charged with 150mL of pineapple extract. Precipitating agent (ethanol, 99.5%) at room temperature is then constantly added at a rate of 0.6mL/s until the final volume of 750mL is reached (Leite et al, 2008). The bulk temperature, represented by TE301, is controlled in order to prevent the enzyme denaturation during the precipitation process.

Setpoint is fixed at 5°C and control action is performed manipulating the coolant flow rate (propylene glycol) by means of pump rotation with the help of a frequency converter. Instrumentation consists of Fieldbus Foundation intelligent devices like temperature and level transmitters. Interface with the system is done by the Distributed Field Interface (DFI302) configured by *Smar Syscon®* software and connected to Scilab through the OPC communication. Process data is monitored with the utilization of display boxes and graphic windows as shown in Figure 3.1 (b).

(a)   (b)
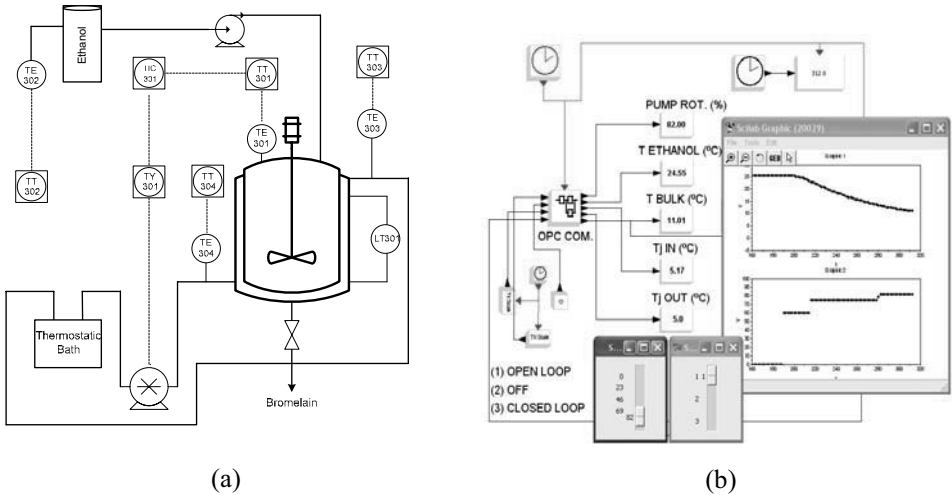
Figure 3.1 Instrumentation diagram and Scicos control application for the bromelain precipitation process.

4. Fuzzy Control

Fuzzy control algorithm is briefly described in the next steps: (i) the inputs error *(E)* and changing of the error *(ΔE)* are weighted in a *fuzzification* process by means of the activation level obtained from their respective membership functions; (ii) influence of inputs on the output is inferred by a set of rules (*IF … THEN… ELSE*) consisting of the linguistic model of the fuzzy controller; (iv) active rules are logically combined and an output fuzzy set is generated; (v) control action *(U)* for the PD (proportional and derivative) algorithm and the changing in control action *(ΔU)* for the PI (proportional and integral) algorithm are obtained from the *defuzzification* of the previously generated output fuzzy set. For a more complete description of fuzzy control from classical PID

approaches see Espinosa and Vandewalle, (1999); Li and Gatland (1995) and Pivoňka (2002).

PI and PD fuzzy control diagrams are shown in Figure 4.1 (a) and (b) respectively. The *FLS* (fuzzy logic structure) block evaluates the fuzzy logic linguistic model. The inputs are scaled before entering the *FLS* by *GE* and *GCE* gains. The outputs are scaled back to crisp values by *GCU* and *GU,* respectively.

(a)   (b)
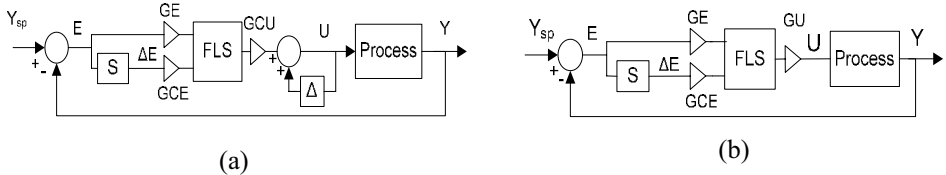
Figure 4.1 Fuzzy logic control diagrams.

The scale gains constitute tuning parameters for the fuzzy controller. For their computation, Pivoňka (2002) methodology was utilized. It relates classical PID parameters *(Kc, Ti, Td)* to the scale gains through a coefficient *M.* This coefficient maps the rule base normalizing the inputs within the limits of [-1, 1]. *Equation (1)* shows the scale gains: τ corresponds to *Ti,* for the PI structure, and *Td* for PD structure.

$$GE = \frac{1}{\tau}; GCE = \frac{\tau}{M}; GU = \frac{KcMTs}{Ti}; GCU = KcM \qquad (1)$$

In this work, inputs *E* and *ΔE* and output *ΔU* consist of seven fuzzy sets (see Figure 4.2): the adjectives are *NB* (negative big), *NM* (negative medium), *NS* (negative small), *ZE* (zero), *PS* (positive small), *PM* (positive medium) and *PB* (positive big). The 7x7 linear rule base for the fuzzy control structures are presented in Table 4.1. Inputs are logically combined by *AND* (product) operator and rules evaluated by *ELSE* (max) operator. The *defuzzification* corresponds to the fuzzy centroid algorithm. PI+PD fuzzy controller corresponds to the parallel combination of PI and PD fuzzy controllers while PI+D fuzzy controller corresponds to the PI fuzzy controller with crisp derivative action.

Table 4.1 Rule base for fuzzy PI/PD controller

e \ Δe	NB	NM	NS	ZE	PS	PM	PB
NB	NB	NB	NB	NM	NS	NS	ZE
NM	NB	NM	NM	NM	NS	ZE	PS
NS	NB	NM	NS	NS	ZE	PS	PM
ZE	NB	NM	NS	ZE	PS	PM	PB
PS	NM	NS	ZE	PS	PS	PM	PB
PM	NS	ZE	PS	PM	PS	PM	PB
PB	ZE	PS	PS	PM	PB	PB	PB
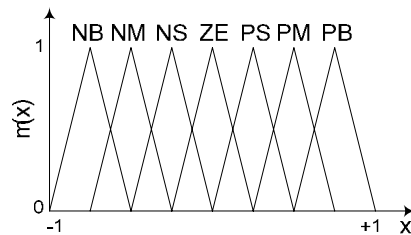
Figure 4.2 Fuzzy set for *E, ΔE* and *ΔU.*

5. Results

The online experimental tests with the proposed control application showed efficient performance for the bromelain precipitation process. Although the sample time errors have not been directly measured, no abnormal situation or errors concerning the system time response were observed. Furthermore, the achieved results are consistent with previous works (see Leite, 2007) performed with the same process.

Figure 5.1 shows the PID controller compared to the open loop response using a fixed rotation of 40% in the coolant pump. The system behavior under PI, PI+D and PI+PD fuzzy controllers is shown in Figure 5.2. Performance indexes are given in Table 5.1. Tuning parameters were investigated by Leite (2007) and consists of $Kc=37\%/°C$ and $Ti=30s$ for the *PI* controller and $Kc=35\%/°C$, $Ti=18s$ and $Td=7s$ for the PID controller. A sampling time of *2s* was utilized. Fuzzy controllers were tuned with the same parameters as the classical PI and PID controllers with a scale coefficient *M=5*.

The large initial overshoot noted in both Figures 5.1 and 5.2 results from the small effective surface area of the wall between the early volumes of precipitating mixture and coolant liquid, heat provided by the dilution of the alcohol in water and by the precipitation process as well.
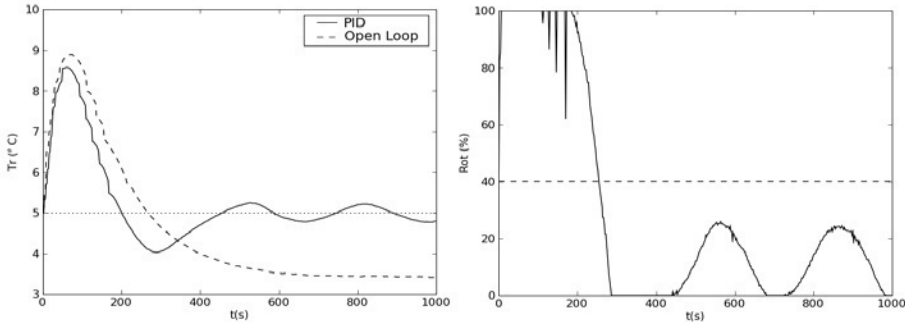
Figure 5.1 System behaviors under classical PID control.
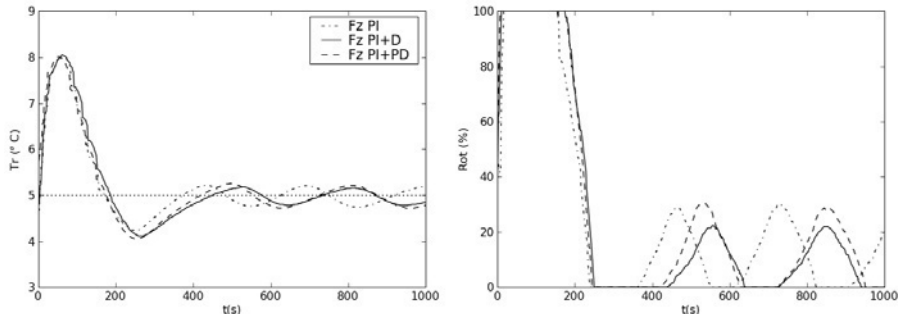
Figure 5.2 System behavior under fuzzy controllers.

As Table 5.1 and Figures 5.1 and 5.2 shows, fuzzy controllers performed better when compared with the classical PID controller. Lower overshoot is a necessary condition for the final product quality. The energy saving, 14% with Fuzzy PI when compared with the PID controller, and the small manipulated variable saturation, required for maintenance cost reduction, are known as key prerequisites for a cost saving production. These characteristics in addition with controller performance index *(ITAE)* indicate the fuzzy PI and the fuzzy PI+D as better control algorithms for the present process.

Table 5.1 Control performance indexes.

Controller	*ITAE*	Overshoot (°C)	MV Saturation (s)	Energy Consumption (kJ)
PID	130424	3.57	184	228
Fuzzy PI	104114	3.05	138	196
Fuzzy PI+D	125430	3.04	162	194
Fuzzy PI+PD	113034	3.01	166	207

6. Conclusions

Through the application of fuzzy control in the bromelain precipitation process, implementation and real-time utilization of advanced controllers in the open source software Scilab were discussed. Results showed better performance of the fuzzy controllers when compared with the classical PID controller. Tracking of control time evaluation confirms Scilab/Scicos suitable for standalone *HMI* and control system in small applications. Future experiments with different control algorithms, such as model predictive control and neural network approaches should be employed in order to promote their utilization not only in the academia, but in industrial applications as well.

References

D. Seborg, T. Edgar & D. Mellichamp, 2004, Process Dynamics and Control, 2nd ed, Wiley.
H. Li & H. Gatland, 1995, A new methodology for designing a fuzzy logic controller, IEEE transactions on systems, man, and cybernetics., 25, 505-512.
J. Espinosa & J. Vandewalle, 1999, Fuzzy Control and PID, Lecture Notes of the Studienamiddaf, Tentoonstelling PID-regelaars: moderene afstellingsmethodes.
M. Leite, 2007, Estudo Comparativo do Desempenho de Controladores Fuzzy e Convencional Aplicados em um Bioprocesso, Universidade Estadual de Campinas, São Paulo - Brasil.
M. Leite, T. Fujiki, F. Silva & A. Fileti, 2008, Determinação de Condições Operacionais de um Processo de Precipitação da Bromelina via Planejamento Experimental, XVII Congresso Brasileiro de Engenharia Química.
M. Najafi, A. Azil & R. Nikoukhah, 2003, Implementation of Continuous-Time Dynamics in Scicos, Proceedings of 15th European Simulation Symposium.
P. Pivoňka, 2002, Comparative Analysis of Fuzzy PI/PD/PID Controller Based on Classical PID Controller Approach, Proceedings of the 2002 IEEE International Conference on Computational Intelligence, 1, 541-546.
R. Nikoukhah & S. Steer, 2006, SCICOS A Dynamic System Builder and Simulator User's Guide, Version 1.0, Institut National de Recherche en Informatique et Automatique.
Y. Shimanuki, 1999, OLE for Process Control (OPC) for New Industrial Automation Systems. IEEE International Conference on Systems, Man, and Cybernetics,6, 1048-1050.
Z. Peng & L. Ma, 2006, The realization of SCADA based on Scilab, Institute of Industrial Control of Zhejiang University.

Flexible and Configurable Steel Plant Scheduling System

Iiro Harjunkoski and Margret Bauer

ABB Corporate Research, Wallstadter Str. 59, 68526 Ladenburg, Germany
E-mail address: iiro.harjunkoski@de.abb.com

Abstract

Steel plant scheduling is very complex, particularly because of the multitude of products and a large variety of possible steel plant configurations and production rules. In this paper, we will present how the steel plant scheduling model based on the one reported in Harjunkoski and Grossmann (2001) has been generalized and adapted to be applied to a real production environment. Besides minor adjustments to the mathematical model, the work involves advanced software development and interface definition as well as implementation and testing.

Keywords: Steel plant, scheduling, connectivity, ISA-95.

1. Introduction

In the steel industry, a strong demand exists for systems that can generate daily production schedules with the assistance of mathematical optimization methods. The demand for such a system stems from two needs: 1) to take the workload off the planning personnel by automatically generating a daily schedule and 2) to potentially increase the production by finding better schedules than is manually possible in the short available time. Depending on the steel plant in question, one need may be more important than the other. In both cases, it is necessary to produce a schedule that reflects the actual operation of the plant and considers all physical production rules. The task is very challenging and numerous, also recent contributions can be found in the literature, for instance Atighehchian et al. (2009) and Pacciarelli and Pranzo (2004).

A steel plant scheduling model describes a multi-stage production process that can schedule different steel grades and families. Considerable effort has been spent on the adaptability of the number and order of stages to allow for all possible plant configurations, see e.g. Harjunkoski and Sand (2008). The model is formulated as a mixed-integer linear programming (MILP) problem and decomposes the overall problem into several sub-problems to achieve fast solution times.

However, even the most carefully constructed production planning method can never see the daylight in a real production environment, unless it can be tied to the real-world situation. This means that the production schedule needs to be generated on-demand and not as a once-off computation. The input data, that is, the list of products to be scheduled, reflects directly the order book and can vary considerable depending on the customer in particular and the market situation in general. Also, plant specific rules may have to be considered that could not be anticipated in a general case, for example, one machine is preferred over another simply because it is newer and more efficient.

In this paper, we will describe the steps and challenges of building a scheduling system. Only the implementation in a real production environment will reveal the validity of any

A Create initial model of problem in GAMS	B Extend and generalize model in GAMS	C Migrate model from GAMS to CPLEX Concert	D Define interfaces (Web Services, ISA-95)	E Compile module (executable)
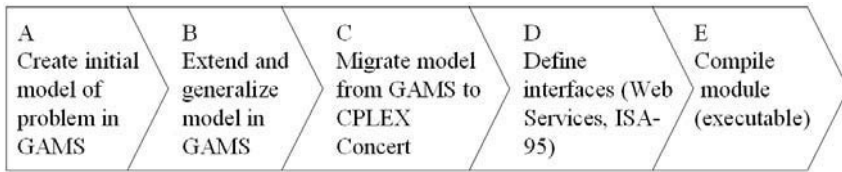
Figure 1. Development steps of a scheduling optimization solution.

model. There is always a certain measure of doubt involved whether the model will work and it is vital to minimize the doubt by generalizing the model for easy adaptation when constructing it. The purpose of this paper is to show the aspects that had to be considered in a real-world implementation and the lessons to be learned for future model developments. These aspects are arguably as important as the derivation of the model. In particular, we will focus on the following:

- *Configurability*: The model has to be adapted to the specific plant configurations which may change throughout the lifetime of the scheduling system.
- *Robustness*: The system must always return an output, at minimum a statement indicating that no result could be obtained.
- *Connectivity*: The system must be able to be integrated into the surrounding computational environment and/or users.
- *User interface*: For industrial application, the critical sets and parameters must be changeable through user friendly interfaces.

In the following sections, these aspects will be discussed in detail. Fig. 1 shows the major development steps that were followed in this work. These steps together with test results from a real-life implementation will be discussed.

2. Implementation aspects

Production in a steel plant has been described in detail in previously (Harjunkoski and Grossmann, 2001). For clarity, the process is briefly introduced in Fig. 2 showing the four stages EAF (electric arc furnace), AOD (argon oxygen decarburization unit), LF (ladle furnace) and CC (continuous caster). As seen in Fig. 2, there can also be more stages and several parallel equipment at each stage and the products (heats) are passed between stages on defined routes.

2.1. Configurability

Ideally, the same model should be applied to different steel plants. A good model should therefore not only focus on one single problem instance, but should be adaptable to various cases. The most important modifications in a steel plant apply to:

- Plant configuration, routing and transportation times
- Certain production rules that may change (setup, clean-up times)
- Maintenance operations
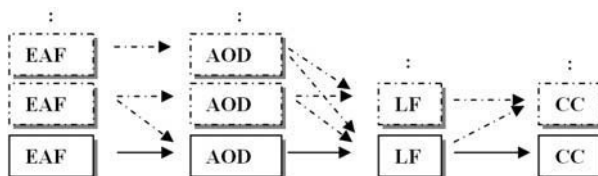
Figure 2. Plant layout and routing between equipment

- Current status of the plant
- Customer / order priority handling (due/release dates, forbidden equipment)
- Objective function components and penalties (lateness, hold-up times, etc)

Also, a steel plant is never static as new process equipment are installed or removed every few years. For example, a new EAF might be added to overcome a bottleneck of steel melting. At the same time, a second equipment may drop out for a long-term maintenance. If the lifespan of the scheduling system should last more than a few months, it is mandatory that it can be easily adapted to incorporate new equipment. If every minor change requires an action of a high-level expert, the engineering costs will be immense. This poses a challenge to the modeling which must be general and adaptable to different plant layouts by changing as few parameters as possible. In Harjunkoski and Sand (2008), the issue of generalizing the scheduling model of a disaggregated solution was discussed. Here, we need to consider all the modeling and solving steps.

Ideally, all issues relating directly to specific production data can simply be handled as parameters in the system input. The model must thus be flexible enough to be able to support various plant layouts, situations in which an equipment is inactive for a given time, changes in transportations times between equipment due to the situation at the site or breakdown of transportation means. In particular, the plant configuration should be handled as model input parameters specifying the number of stages and equipment.

Basic production data can be entered as standard model parameters, e.g. production times and due dates, which are typically represented in any scheduling model. This becomes more complex when production times vary. Production times then have to be expressed as a variable in the model with lower and upper bounds. This may make some production constraints non-linear and complicates significantly the use of a discrete-time representation. Therefore, this work is restricted to a continuous-time approach.

2.2. Robustness

During implementation, different use cases of the scheduling system have to be defined, that is, generation of for example daily or weekly schedules that reveal different characteristics of the production environment. Some products may trigger certain rules, others not and some may not even be captured by the implementation. Scheduling twenty products of the same type is certainly a different task than scheduling twenty different products. The use cases thus result in different problem size and complexity.

When supporting the actual running of a production, the scheduling system should always produce a result – and if not possible this should be clearly communicated to the user with some hints of a possible problem source. Also, the solution time has to be within reasonable limits for all the use cases. It is usually not acceptable to let the user
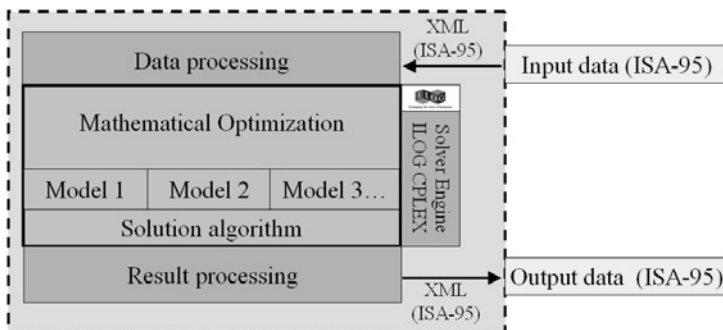
Figure 3. System architectural overview.

ProdSchedule
  ProductionRequest
    ID
    Description
    StartTime
    EndTime
    ProductionParameter
    SegmentRequirement
      ID
        EAF
      ProductSegmentID
      Description
      EarliestStartTime
      LatestEndTime
      Duration
      ProductionParameter
      ProductionParameter
      EquipmentRequirement
        EquipmentID
        Description
        EquipmentRequirementProperty
        EquipmentRequirementProperty
        EquipmentRequirementProperty
      EquipmentRequirement
    SegmentRequirement
    SegmentRequirement
    SegmentRequirement
  ProductionRequest
  ProductionRequest

Production Request	
Fieldname	**Fieldname desription**
ID	Unique ID of production request (order)
Description	Description of the production request, for example steel grade
StartTime	Scheduled start time of production request
EndTime	Scheduled end time of production request
ProductionParameter	Parameter related to the order
SegmentRequirement	Requirements of the segments or stages (e.g. EAF) to which the production requests may be assigned
EquipmentRequirement	Requirements of the equipments (e.g. EAF 1) to which the production requests may be assigned

Fig 4. ISA-95 XML structure highlighting a production schedule and description of fieldnames.

wait for the result for more than a couple of minutes. The optimization routine should be aborted after an agreed time, returning the message to possibly reduce the number of products. Developing a system that can return an intelligent result in all (foreseeable) situations requires actions on all levels: from the modeling to the software development. The approach here varies strongly from the commonly applied scheduling systems in that respect that we aim to optimality or close-to-optimality, i.e. utilize MILP methods as far as possible, supported by intelligent heuristics that makes the solution several magnitudes faster without losing the strive to optimality.

For instance the disaggregation process in Harjunkoski and Grossmann (2001) builds upon the idea of splitting and merging large optimization problems by the use of optimization methods. The various models can easily be covered by the solution structure, as seen in Fig. 3. Other possibilities includes using time limits and if a solution should be available at the given time the optimization should simply wait until a first feasible solution has been found or abort the task and inform about the limiting time constraint.

A common challenge for real production problems is that as long as a system can handle a certain amount of order efficiently, the user naturally wants to increase the number of orders. This leads to a continuously growing problem size and the fact of combinatorial explosion must be considered in the core of the solution. It is therefore necessary to agree on the outset on the time window, that is, daily or weekly schedule, and thus the maximum number of orders that can be scheduled.

2.3. Connectivity

The scheduling system should be embedded in the IT architecture of the steel plant. This concerns the inputs and the outputs as well as the parameterization of the system. The input data comprises the production orders as well as the plant configuration parameters. As a result, the model is decoupled from the graphical user interface and the visualization development work can be carried out separately from the model adaptations. In order to enable a highly flexible and well defined connection, ISA-95

compliant interfaces through XML files were defined. The ISA-95 standard describes models and terminology for the information exchange between enterprise resource planning (ERP) and, among others, scheduling system. Fig. 4 shows an example of an ISA-95 structure. Both production orders and production history are normally entered and stored in the ERP. ISA-95 as a standard does not provide all the needed elements for such a specific industry and problem. However, these have been added in a way that is enabled by the standard.

Fig. 3 shows that the optimization problem itself is implemented in a very flexible way. The required optimization models can be added or modified and are coordinated by a solution algorithm. The solution algorithm calls the optimization models as needed, using the underlying MILP/LP solver ILOG CPLEX Concert Technology.

Data pre-processing is mainly required for consistency checking of the input data and for converting it into the solution memory data structures. For example, it is easy to capture the current status incorrectly and assign two products to a machine at the start. This has to be checked before starting the optimization and a message with an explanatory error message should be returned.

The output XML file contains the start and end time of each product on each equipment and is also ISA-95 compliant as it may be stored in the ERP system. The result post-processing is necessary to generate the output file from the optimization results consistently with the input data.

2.4. User Interface

The user interface has to accommodate for two main tasks as highlighted in Table 1: Generate a production schedule and configure the production parameters. Configuration is a less frequent task that should be only carried out by more senior users with an overview insight of the complete production process. The standard functionality is to select an input file, start the optimization and then display the results. The input file is a list of products specifying the product ID, type, due date and current production status. When generating a rolling schedule, some parts of the products are already in the process and it is necessary to capture which product is already at what stage on which machine. The output is a list of these products specifying the product ID, start time and end time on each equipment. The output is best displayed in a Gantt chart.

Arguably the biggest challenge when designing a scheduling system is to accommodate for the needs of the person who operates that system. Any intelligent person operating the system will only trust and then use it if she a) understands why the schedule has been generated in the way it has, b) tune the behaviour and c) if she has the possibility to manually change the automatically generated schedule. The latter reason is a much debated issue.

Task description	Actions	User	Frequency
Generate production schedule	• Select input file for new schedule • Start optimization • Display the generated schedule	Scheduler	Several times a day
Configure production parameters	• Specify the production / operational parameter • Upload changes	Plant operating manager	Infrequent

Table 1. Description of use cases of the graphical user interface.
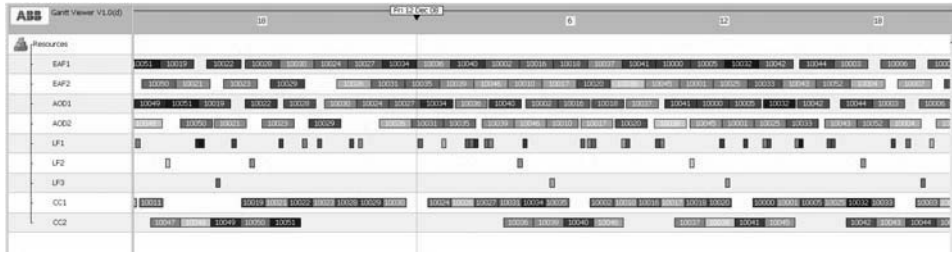
Fig 5. A Gantt Chart of an implemented test case solution

3. System Development Steps and Test Cases

The more the implementation aspects of configurability, user interaction, robustness and connectivity have been considered in conceptual modeling phase, the easier, the shorter the system development will be. Steps for the model adaptation are as follows:

1. Collect production rules in interviews
2. Define use case scenarios and desired results
3. Iteratively
   a. Test the use case scenarios
   b. Discuss results
   c. Adapt model to consider rules

It is important to realize that the model cannot be adapted in one single shot and the recursion has to be repeated several times. Also, it is quite likely that the use cases will not cover all situations that may later occur when the system is in operation over a number of years. Thus, exceptions have to be anticipated in all variables and parameters.

The system can generate acceptable solution to normal two to three days scheduling problems within a few seconds / minutes. An example of a resulting Gantt Chart with around 50 heats and nine equipment is shown in Fig. 5.

4. Conclusions

In this paper, we have discussed the challenges that occur when developing a scheduling solution from an existing scheduling model. Only in very rare, possibly never heard of, cases can the model be used without any changes. If you have constructed a scheduling model then it is most likely that it requires adaptation to the specific plant configuration. The more generic and flexible the model has been formulated with an industrial environment in mind, the easier the development of the scheduling system.

References

Atighehchian A., Bijari M. and Tarkesh H. (2009). A novel hybrid algorithm for scheduling steel-making continuous casting production, Computers & Operations Research, Volume 36, Issue 8, August 2009, pp. 2450-2461

Harjunkoski I. and Grossmann I.E. (2001). A Decomposition Approach for the Scheduling of a Steel Plant Production, Computers chem. Engng, 25, pp. 1647-1660

Harjunkoski I. and Sand G. (2008). Flexible and Configurable MILP-Models for Meltshop Scheduling Optimization. Proceedings, 18th European Symposium on Computer Aided Process Engineering, Lyon, June 1-4, 2008

Pacciarelli D. and Pranzo M. (2004). Production scheduling in a steelmaking-continuous casting plant. Computers & Chemical Engineering, Volume 28, Issue 12, pp. 2823-2835

Monitoring Penicillin G Acylase (PGA) Production Using Principal Component Analysis (PCA)

Edson R. Nucci, Antonio J.G. Cruz, Raquel L.C. Giordano and Roberto C. Giordano

LaDABIO – Laboratory for Development and Automation of Bioprocesses, Department of Chemical Engineering, Federal University of São Carlos, P.O. Box 676, 13565–905, São Carlos – SP, Brazil. E-mail address: roberto@ufscar.br

Abstract

The complexity of biological processes turns infeasible the development of detailed, structured phenomenological models of the cultivation of microorganisms in bioreactors. Therefore, cause-effect relations between on-line measurements and the state variables that are important for the optimal operation of industrial fermenters are sometimes hard to ascertain. In this context, data pre-treatment techniques are useful for control and fault detection. Among them, principal component analysis (PCA) has an important role. This work presents a case study of the application of this technique during real experiments, where the enzyme penicillin G acylase (PGA) was being produced by *Bacillus megaterium*. PGA hydrolyzes penicillin G to yield 6-aminopenicilanic acid (6-APA) and phenyl acetic acid. 6-APA is an important substrate, used to produce semi-synthetic β-lactam antibiotics. The algorithm was implemented for on-line detection of deviations from the desired process behavior. The experiments were carried out in a 2-liter bioreactor (Applikon®) operated in batch and fed-batch mode in different experimental conditions. A static PCA model was initially developed for the process, and its results are presented here. Hotteling's T^2 was the discrimination criterion employed in this multivariable problem, and the method showed a high sensibility for fault detection in all real cases that were studied.

Keywords: Multivariate statistical process control (MSPC), fault detection, *Bacillus megaterium* cultivation, bioreactor monitoring

1. Introduction

Process monitoring and fault diagnosis are the determinant for the successful operation of bioreactors and for quality control. In batch and fed-batch processes, small changes in the operating conditions may impact severely the quality of the desired product, frequently a bio-molecule which has its concentration analyzed off-line. If the product does not complain to the quality standards, the entire batch is lost. The industrial demand for reliable methodologies for on-line fault detection is therefore evident. The implementation of early fault detection algorithms would allow the operator to take corrective measures before the batch process is accomplished (Xu et al., 2006).

A large amount of process data can be collect and stored in modern biotechnology industries, whose processes have computer-based monitoring and control. Therefore, multivariate statistical process control (MSPC) has been widely used for fermentation processes (Hu and Yuan, 2008; Gnoth *et al.*, 2008). In MSPC the correlation between the original variables is considered, thus decreasing the number of false alarms. Statistical tools such as principal component analysis (PCA) are commonly used to explore historical data (Lopes and Menezes, 2004). Most often, PCA is employed in

statistical control to get a quick overview and to detect deviations from the desired process behavior (Chiang *et al.*, 2006; Al-Alawi *et al.*, 2008; Gnoth *et al.*, 2008).

This work presents a real case study of the on-line application of PCA in order to recognize fluctuations and their respective effects on faults that occurred during the cultication of *B. megaterium* for the production of the enzyme penicillin G acylase (PGA). Penicillin G acylase (E.C.3.5.1.11) is an enzyme of great importance for the pharmaceutical industry, used in the production of 6-aminopenicilanic acid (6-APA), a key component for the synthesis of β-lactam antibiotics.

2. Material and Methods

The microorganism employed was *Bacillus megaterium* ATCC 14945, from Fundação Tropical (Campinas, SP, Brazil). The experiments were carried out in a 2-liter bioreactor (Applikon®). Dissolved oxygen concentration was controlled by manipulating stirrer speed, temperature was kept at 30°C and the aeration rate at 2.0 L.min^{-1} (1 atm and 21°C). The bioreactor was coupled to a data acquisition system (National Instruments®) and on-line variables were stored each 10 seconds. Further details about the composition, procedure and laboratory analyses can be found in Silva et al. (2006) and Nucci et al. (2007).

3. Methodology

PCA is used to capture the major normal statistical correlation between measurements of process variables, from residuals that quantify their main variances. Some kind of statistics of these residuals can then be used to detect and diagnose faults (Xu et al., 2007). PCA involves finding the eigenvalues of the sample covariance matrix, which are the variances of the principal components. For a normalized (mean centered, variance scaled) sample Matrix *X[n,m]* with *n* samples and *m* variables, PCA will find *m* uncorrelated new variables, the variance of which decreases from first to last (Albert and Kinley, 2001). Let the new variables be represented by t_i for a particular sample *i* as follows:

$$t_i = \sum_{j=1}^{m} X_j \times p_{ji} \tag{1}$$

The first principal component t_1 is found by selecting p_i, so that t_1 has the largest possible variance subject to the condition shown by Equation 2:

$$\sum_{i=1}^{m} p_i^2 = 1 \tag{2}$$

In covariance matrix, c_{ij} is the covariance between variables X_i and X_j. Of course, the diagonal element c_{ii} is the variance of X_i. The variances of the individual principal components are the eigenvalues of the matrix *C*, and the sum of the eigenvalues is equal to the sum of the variances of the original variables. For *m* input variables there will be *m* principal components, some of which might be negligible, if the original variables were either correlated or collinear. By retaining only the first *r* principal components, the *X* matrix is approximated by the following equation:

$$\hat{X} = \sum_{i=1}^{r} t_i \times p_i^T + E \tag{3}$$

In equation 3, *E* is the residual matrix, *p[m,n]* are the loading and *t[n,r]* are the scores.

This transformation results in several desirable mathematical and statistical properties that are associated with the transformed data (scores), enabling the derivation of statistical confidence limits. If the original variables are correlated, a reduced number of control charts

can be achieved ($r<m$) (Albert and Kinley, 2001). The ability to adequately represent an m variable data set in $r<m$ dimensions is one the main advantages of PCA, that is, data compression. More details about T^2 Hotelling, control limits and contribution plot can be found in AlGhazzawi and Lennox (2007); Hu and Yuan (2008).

4. Results and Discussion

A data set of seven runs of the PGA production process was used to test the procedure, with 7291 samples in normal operation conditions. 4448 samples were used for testing and validation. The monitored variables are listed in Table 1. The average length of cultivation is 24 h, and the sampling interval for laboratorial analyses is 2 h

Table 1 Variables monitored during penicillin G acylase production

No.	*Process variables*
1	agitation speed, rpm
2	aeration rate, $L.min^{-1}$
3	concentration of dissolved oxygen in the fermentation media, % saturation
4	pH
5	temperature, °C
6	fraction molar of CO_2, %

Fig. 1 shows T^2 analysis of a faultless experiment. The chart shows no violation of the control limits during all experiment, except for the initial samples, corresponding to the start-up of the process.
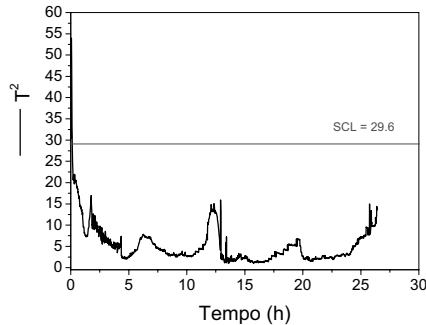
Figure 1 – On-line monitoring chart without faults. SCL = superior control limit.

Figure 2 shows a run where a fault occurred, the aeration rate (corresponding to variable 2) decreased from 2 $L.min^{-1}$ (the set-point) to 1.4 $L.min^{-1}$ approximately after 3 h of cultivation, see Figure 2(a). Figure 2(b) shows the sensitivity of the algorithm. Around 2 h the threshold of alarm (SCL = 29.6) was almost reached, and at 3 h there is a clear detection of the fault. The aeration rate influences the supplement of oxygen to the strictly aerobic microorganism: the decrease of oxygen mass transfer from the gas phase to the liquid medium, due to the smaller velocity of the air stream, will diminish the dissolved oxygen concentration thus restraining the biomass growth. It should be stressed that the operator might well have missed the deterioration of the air flow control in a normal situation, and so the chart (Fig. 2b) proves to be an important tool, indicating the occurrence of a fault.
Fig. 3 shows the contributions to the fault of each process variable. As expected, variable 2 provided the largest contribution to the deviation of the process from the normal operation region. Hence, the aeration rate is identified to be the source of the fault. Earlier detection of the faults in batch process and determination of its root cause will guide the operator to prompt corrective measures.

(a)   (b)
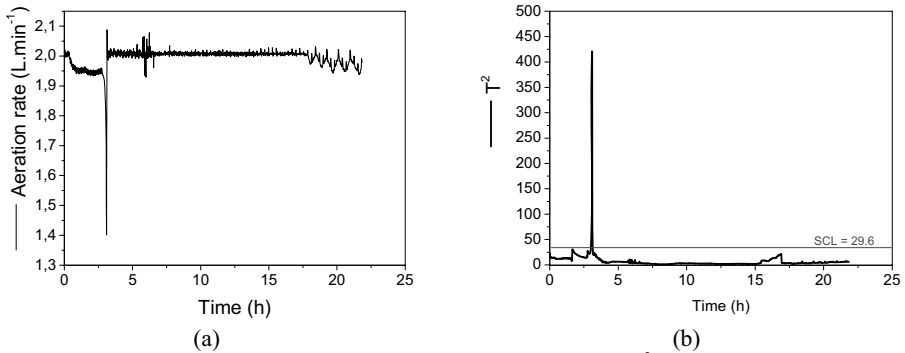
Figure 2 – (a) On-line monitoring chart for fault 1 (aeration); (b) T^2 on-line monitoring charts using PCA for fault 1
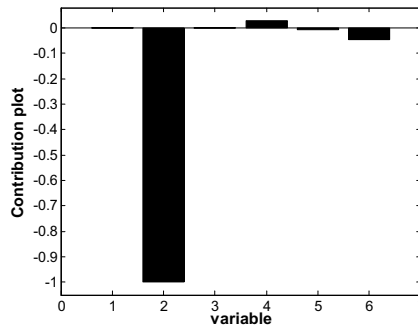
Figure 3 – Contribution plot for the aeration rate fault. Variable 1.Agitation speed; 2. Aeration rate; 3.Dissolved oxygen; 4.pH; 5. Temperature; 6.Fraction molar of CO_2.

Robustness is an important characteristic of monitoring and fault detection. In almost every real industrial process, measured variables inherently contain noise that originates from various sources such as measurement devices and electrical equipment. Besides, outliers usually exist in sample data. If the signal-to-noise ratio is small, one may encounter misleading or biased results at the stages of data processing (Hu and Yuan, 2008).

In the previous fault, the input variable that caused the abnormal operation was directly monitored. In order to test the robustness of the method, this variable was removed from the samples data set, and the method was once again applied, now off-line, with the purpose of testing its robustness. Two other variables were introduced in this test. These variables were not used formerly, though being available in the data acquisition system. They are: the molar fraction of O_2 at the exhaust gas and the respiratory quotient (RQ), which is the ratio between produced CO_2 and consumed O_2 (calculated from the overall mass balance).

In this way, a new situation was emulated: instead of directly measuring the input variable responsible for the fault (the air flow rate), an output operational variable was sampled (molar fraction of oxygen at the gas outlet), and a second one was calculated (the RQ). Of course, there is a cause-effect correlation between the inlet air flow, the metabolic state of the culture and, consequently, the O_2 in the effluent gas and the RQ. Figure 4 shows that the method was able to identify the fault correctly, also in this case.

It should be noticed that this is a common situation in real processes, i.e., a non-monitored input variable reaching an abnormal level and causing a fault. In our case, the output measurements were able to provide enough information for the PCA algorithm to identify the fault. Fig 4a shows violations of the control limit between 2 and 3 h. The

contribution plot (Fig. 4b) indicated a problem with variables 5, 6 and 7 (molar fraction of CO_2, of O_2 and RQ, respectively), i.e., these variable are strongly correlated. These results illustrate the robustness and efficiency of the method. Indeed, these variables are correlated, and directly affected by the reduction of the inlet flow rate of air.

<center>(a)</center> <center>(b)</center>
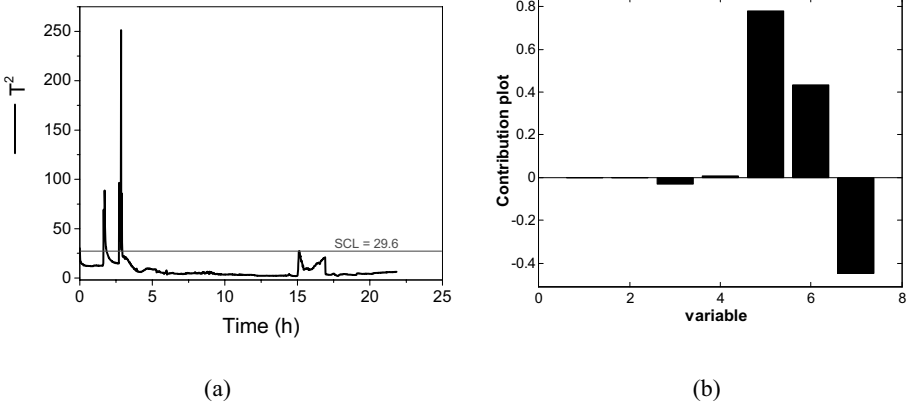
Figure 4– (a) T^2 on-line monitoring charts using PCA for fault 3; (b) Contribution plot for the pH fault. Variable 1.Agitation speed; 2.Dissolved oxygen; 3.pH; 4. Temperature; 5.Fraction molar of CO_2; 6.Fraction molar of O_2; 7. Respiratory quotient (RQ)

For the run shown in Figure 5, two faults occurred. The pH decreased until 6 h (after this time it started to increase from 6.95 to 7.30). And a problem appeared during the aeration rate data acquisition. The pH influences the growth of the microorganism, and consequently the enzyme activity because the growth is associated with the production of the enzyme. Figure 5a shows the T^2 on-line monitoring chart. It can be seen that between 12 and 14 h there is a violation of the control limit (SCL = 29.7). Fig. 5b shows the contributions of each of the process variables. This chart correctly highlights two variables as the source of the problem, variables 2 and 4

<center>(b)</center>

<center>(a)</center>
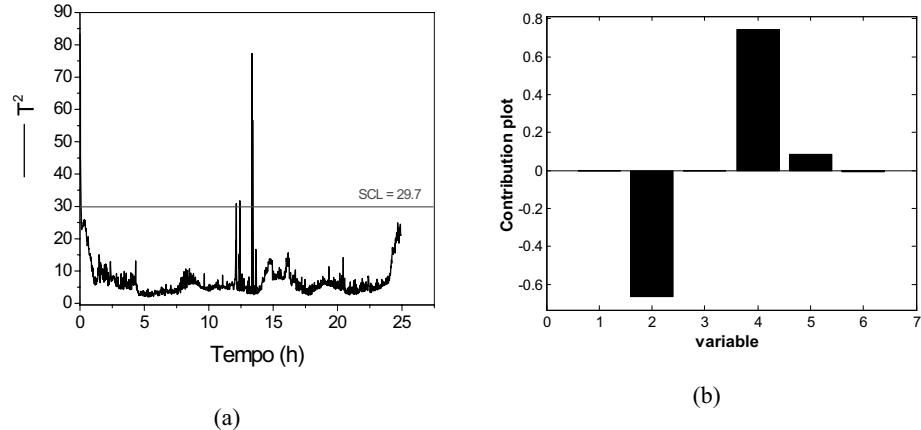
Figure 5 – (a) T^2 on-line monitoring charts using PCA for fault 2; (b) Contribution plot for the pH and aeration rate fault. Variable 1.Agitation speed; 2. Aeration rate; 3.Dissolved oxygen; 4.pH; 5. Temperature; 6. Molar fraction of CO_2

In this case, besides an input variable, the air flow rate, a state variable, the pH, was also at abnormal values. These two variables were removed from the data set and the PCA was re-calculated with two new process variables, i.e., molar fraction of O_2 and RQ. T^2 chart was

calculated and similar results were obtained as Fig. 5a. On the other hand, the contribution plot, indicates those variables 4, 5 and 6 (fraction molar of CO_2, fraction molar of O_2 and RQ, respectively) provide the largest contributions to the deviation of the process from the normal operational region. Once again, the variables more closely related to the metabolic state of the culture had the higher contribution. It should be stressed that the term contribution here is within a statistical framework, and does not reflect a cause-effect relation. Anyway, the operator should translate these results as an indication that the deviation from the acceptable operational region is causing a metabolic stress to the microorganism, and thus seek for the causes of this unexpected behavior, but already excluding the variables with almost null contribution. Hu and Yuan (2008) obtained similar results using PCA during a cephalosporin fed-batch fermentation process. The authors indentified the faults in two abnormal experiments (decrease in aeration rate and agitation speed).

5. Conclusion

This article applied PCA (*principal component analysis*) to a bioreactor producing an enzyme, PGA, growing the strictly aerobic bacterium *B. megaterium* in an agitated and aerated tank. The method was able to filter normal random oscillations, and could detect faults without false positives. The obtained results demonstrate the power and advantages of the proposed approach. It was efficient and robust, identifying the abnormal data during the real experiments of validation in all cases. Even when the variable that caused the fault was not monitored, the results were good enough to support decision-making and the intervention of the operator. Therefore, this method may be very useful to ensure that each production run is a close as possible to pre-established optimal process trajectories.

Acknowledgments

References
Albert, S., Kinley, R.D., 2001, Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. TRENDS in Biotechnology, 19(2), 53-62.

AlGazzawi, A., Lennox, B., 2008, Monitoring a complex refining process using multivariate statistics. Control Engineering Practice,.16, 294-307.

Al-Alawi, S.M., Abdul-Wahab, S.A., Bakheit, C.S., 2008, Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. Environmental Modelling & Software. 23, 396-403.

Chiang, L.H., Leardi, R., Pell, R.J., Seasholtz, M.B., 2006, Industrial experiences with multivariate statistical analysis of batch process data. Chemometrics and intelligent laboratory systems. 81, 109-119.

Gonth, S., Jenzsch, M., Simutis, R., Lubbert, A., 2008, Control of cultivation processes for recombinant protein production: a review. Bioprocess and Biosystem Engineering. 31, 21-39.

Hu, K., Yuan, J., 2008, Statistical monitoring of fed-batch process using dynamic multiway neighborhood preserving embedding. . Chemometrics and intelligent laboratory systems. 90, 195-203.

Lopes, J.A., Menezes, J.C., 2004, Multivariate monitoring of fermentation process with non-linear modelling methods. Analytica Chimica Acta. 515, 101-108.

Nucci, E.R., Silva, R.G., Souza, V.R., Giordano, R.L.C., Giordano, R.C., Cruz, A.J.G., 2007, Comparing the performance of multi layer perceptrons networks and neuro-fuzzy systems for on-line inference of *Bacillus megaterium* cellular concentrations. Bioprocess and Biosystem Engineering, 30, 429-438.

Silva, R.G., Souza, V.R. Nucci, E.R., Pinotti, L.M., Cruz, A.J.G, Giordano, R.C., Giordano, R.L.C., 2006, Using a medium of free amino acids to produce penicillin G acylase in fed-batch cultivations of *Bacillus megaterium* ATCC 14945. Brazilian Journal of Chemical Engineering, 23, 37-43.

Xu, Z., Weiwu, Y., Huihe, S., 2006, Monitoring and Fault Diagnosis for Batch Process in Feature Extract in Fisher Subspace. Chinese J. chem. Eng., 14(6), 759-764.

Stochastic and Deterministic Performance Assessment of PID and MPC Controllers: Application to a Hydrotreater Reactor

Alain C. Carelli, Mauricio B. de Souza Jr.

Department of Chemical Engineering – Federal University of Rio de Janeiro
Av. Athos da Silveira Ramos, n. 149 - CT - Bloco E, Sala E-209, CEP 21941-909,
Rio de Janeiro – RJ, Brazil

Abstract

Hydrotreating processes are very important in petroleum refineries due to environmental reasons. However, the hydrotreater operational costs are very high, being increased when there is not an optimization on the catalyst lifetime, which is harmed by an inefficient control. At the same time, a control system tends to lose performance over time if its response is not monitored and thus there is no support information on how to adjust it. Enhanced controller projects have complementary systems to identify and diagnose performance reduction and also to implement predetermined solutions vis-à-vis the desirable type of output. When selecting the controller design and strategy, the thorough simulation of the underlying process is important in order to draw a greater number of possible alternatives of control and their application on this simulator, thus conferring more information for the decision. The goal of this paper was to investigate the response of feedback controllers in contrast with model predictive ones applied to the first bed of a diesel hydrotreating reactor, which was simulated using a phenomenological model, and to asset their deterministic and stochastic performances. The deterministic performance was evaluated through the integral of time multiplied by the absolute error (ITAE) and the quadratic variation of the control action; the stochastic performance was analyzed based on the normalized Harris index and on the variability and quadratic variation of the control actions. The predictive control algorithm used in this paper was the Generalized Predictive Control (GPC) one, developed by Clarke *et al.* in 1987. Besides comparing different controller designs, this work also analyzed the impact of the changes in the tuning parameters. The feedback controllers were tuned by different heuristic and model based methods. The relation between predictive control performance and variations in reference trajectory and internal model was investigated. This study adds background information on deterministic and stochastic performances of process controllers, through the application of different designs and strategies of controllers to a hydrotreating complex reactional process.

Keywords: process control, model predictive control, controller audit, petroleum refinery.

1. Introduction

Global environmental concern has pushed the contaminant contents in fuel oil specifications to lower levels, stimulating conscious investments in petroleum refineries with the goal of guaranteeing larger removal of contaminants, such as sulfur and nitrogen. Hydrotreating units are being improved and developed with new process control technologies, allowing cleaner output of fuel oils through reactions as

hydrodesulfurization, hydrodenitrogenation, hydrodeoxygenation, hydrodearomatization, hydrodemetallization and hydrodeasphaltenization.

Modern industries present many automatic control loops making the use of assessment tools mandatory in order to properly monitor all of them. MITCHELL and SHOOK (p. 3 – 4, 2005) related their experience in Model Predictive Control, showing four paths that MPC control applications may follow, depending on the level of support and monitoring they receive: 25 % fail within just months of commissioning; 40 % fail approximately 3 years of commissioning with inadequate support and no monitoring; 25 % continue to deliver benefits – albeit at a reduced level – with good support and no monitoring and 10 % will deliver increased benefits over time with adequate support and online monitoring.

The main aim of this paper is to compare the deterministic and stochastic performances of feedback and predictive controllers applied to a diesel hydrotreating reactor which was simulated using a phenomenological model.

2. Conceptual Aspects

2.1. Diesel Hydrotreating

The hydrotreating (HDT) unit considered in this work employs two trickle bed reactors (TBR) in series, each composed by two fixed beds, as showed in Figure 1. The oil feed is combined with makeup and recycle hydrogen and heated to the reactor inlet temperature. Heat is provided by heat exchange with the reactor effluent and by a pre-heater furnace. The reaction of hydrogen and oil occurs in the reactors in the presence of the catalyst. Quench gas (cold recycled hydrogen gas) is added between reactors and between catalyst beds of the reactors in order to maintain reactor temperatures in the desired range. The 2nd reactor effluent exchanges heat with the reactor feed in order to recover the heat released from the reactions. After cooling, the reactor effluent is flashed in the hot, high-pressure separator (HHPS) to recover hydrogen and to make a rough split between light and heavy reaction products. The liquid from HHPS has its pressure lowered, than it is sent to the low-pressure separators, and on to the product fractionator. The HHPS vapor is cooled and water is injected to absorb hydrogen sulfide and ammonia produced in the reactors. The mixture is further cooled to condense naphtha and gas oil and is flashed in the cold, high-pressure separator (CHPS). The CHPS separates vapor, liquid water, and liquid light hydrocarbons. The pressure of the liquid hydrocarbon is lowered and it is sent to the low-pressure separators. The water is sent to a sour water recovery unit for removal of the hydrogen sulfide and ammonia. The hydrogen-rich gas from the CHPS flows to the H_2S absorber. The purified gas flows to the recycle compressor where it is increased in pressure so that it can be used as quench gas and recombined with the feed oil. Liquid from the low-pressure separators is fed to the atmospheric fractionator, which splits the hydroprocessed oil from the reactors into the desired final products.

The model adopted here to represent the HDT reactor was introduced by CARNEIRO (1992) and applies the concept proposed by HLAVÁCEK (1982) to represent a fixed bed reactor through multiple CSTR-CELLs in series. The CSTR-CELL model considers mass and heat axial dispersion in the bed, mass diffusion and heat transportation between fluid and solid phases. The following assumptions were made: only one 1st order – with respect to the mean concentration of a pseudo-reagent A in the solid phase porous – reaction occurs and the reaction rate can be described by the Arrhenius equation; there is no volume variation in the reactor; the reactors are adiabatic; there is only one liquid and one solid phase with constant physical-chemical properties; there is

only longitudinal transport phenomena; and, there are non-linear interactions between kinetic and thermal processes.
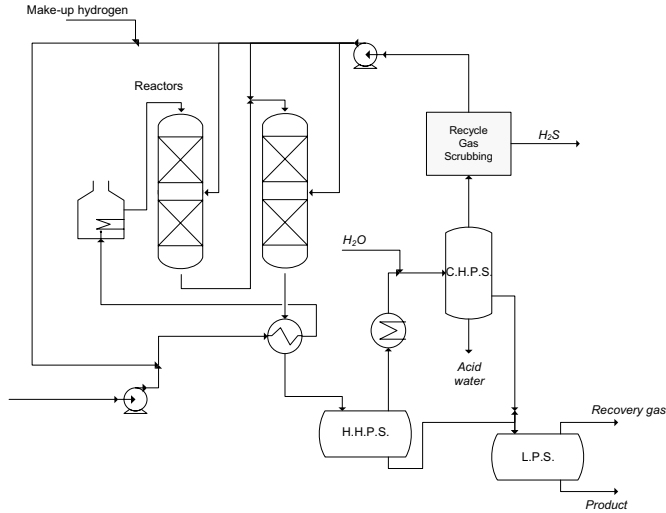
Figure 1 – Diesel hydrotreating process

CARNEIRO's (1992) model was employed here for being both able to represent the main process dynamics and simple, as it is composed only by ordinary differential equations.

2.2. Model Predictive Control – Generalized Predictive Control

The Generalized Predictive Control (GPC) algorithm covers a large variety of control goals in contrast to other methods, so that some of them can even be considered as GPC specific cases (CLARKE *et al.*, 1987).

With the premise that all process natural disturbances can be characterized by a stochastic disturbance, the principle of the superposition can be used to represent all disturbances as a unique influence in the output. Then, the process can be described by the following CARIMA (controlled auto-regressive and integrated moving average) model:

$$A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t-1) + C(q^{-1})e(t)/\Delta(q^{-1}) \tag{1}$$

where A, B and *C* are polynomial in the backward shift operator q^{-1}, *d* is the dead-time, *e(t)* is an uncorrelated random sequence and $\Delta(q^{-1})$ is the differencing operator $1 - q^{-1}$.

The CARIMA model may be considered to be the most appropriated model for many industrial applications with non-stationary disturbances. In practice, it has two main types of disturbance: occurrence of random steps in random interval (e.g. changing of the product quality) and Brownian movement which is met in plants that depend on the energy balance (CLARKE, 1988).

The following cost function is assumed:

$$J = (Gu + f - w)^T (Gu + f - w) + \lambda u^T u \tag{2}$$

where, $G(q^{-1}) = E(q^{-1})B(q^{-1})$; $E(q^{-1})$ and *f* are provided by the Diophantine Equation; *w* is the reference trajectory or *set-point*; and, λ is a weighting sequence.

Assuming that there are no constraints in the control actions, the minimum of *J* can be met by equating to zero the gradient of the cost function with respect to the control actions. Therefore, the following result is used in order to obtain the control actions:

$$\Delta u = (G^T G + \lambda I)^{-1} G^T (w - f) \tag{3}$$

3. Methodology

The application of the controllers was not proceeded for all the control loops in the hydrotreating unit because the goal here was the comparison among several controller performances. The controller considered in this work was the master one that controls the outlet temperature of the first bed in the first reactor while manipulating the set-point for the slave controller. The slave controller controls the inlet temperature of this bed by manipulating the fuel flow that enters in the pre-heater furnace. However, the slave controller was disregarded of this work due to the insignificant time constant of the slave loop in contrast to the master loop.

3.1. Tuning Feedback Control

The feedback controller in the referred control loop was tuned by four different methods: Ziegler & Nichols (*ZN*), Cohen & Coon (*CC*), Error Integral (*EI*) and Internal Model Controller (*IMC*). Considering proportional (*P*), proportional-integral (*PI*) and proportional-integral-derivative (*PID*) controllers, ten feedback controllers (P/PI/PID – ZN; P/PI/PID – CC; PI/PID – EI; PI/PID – IMC) were tested. In Table 1, the tuning parameter values of each feedback controller can be observed.

Table1: Tuning feedback parameters

METHODS	Controller Type	Proportional term (K)	Integral term (τ_I)	Derivative term (τ_D)
Ziegler & Nichols	P	0.58	-	-
	PI	0.52	131.53	-
	PID	0.70	78.92	19.73
Cohen & Coon	P	0.72	-	-
	PI	0.52	272.75	-
	PID	0.84	953.22	160.96
Error Integral	PI	0.46	776.25	-
	PID	0.72	367.64	193.64
IMC	PI	0.30	557	-
	PID	0.60	811	174.45

3.2. Predictive Control Design

The controller internal model is the function which mimics the process and influences all the control strategy deduction. Two models were investigated here: one – given by Equation (4) – which describes the simulated process data very well and another – represented by Equation (5) – which renders an inferior description of the simulated process data.

$$\frac{y}{u} = \frac{2.06}{557s+1}e^{-500s} \tag{4}$$

$$\frac{y}{u} = \frac{2.06}{300s+1}e^{-600s} \tag{5}$$

The function considered to represent the reference trajectory was a first order model, with an adjustable parameter α. The larger α, the more cautious are the control actions and, inversely, the smaller α, the more energetic they behave. If α is equal to zero, the trajectory is constant and equal to the set-point. Three cases were analyzed by varying the internal model and α: *I.* good internal model (4) with α equal to 0.8; *II.* worse internal model (5) with α equal to 0.8; and, *III.* good internal model (4) with α equal to 0.7. The prediction (*N*) and control (*NU*) horizons were held, respectively, equal to 4 and 3 sampling times for the cases 1 and 3 and equal to 8 and 3 sampling times for the case 2. These parameters were adjusted because of change in the internal model.

3.3. Controller Performance Assessment

The deterministic regulatory performance assessment was accomplished through two indexes: the integral of the product between the time and the absolute error (*ITAE*), to evaluate the output behavior, and the mean of the squared variations in the control actions (*Su*), to evaluate the controller aggressiveness. The *Su* can be calculated through the following equation:

$$Su = \frac{1}{n+1}\sum_{k=0}^{n}\Delta u^2(k) \tag{6}$$

where n in the number of available data points. The following disturbance equation was introduced in the controlled variable for the stochastic performance assessment:

$$C(q^{-1})e(t) = e(t) + 0.5e(t-1) + 0.25e(t-2) \tag{7}$$

Three indexes were calculated to assess the stochastic performance of the controllers: the normalized performance index (ANORM), only for the feedback controllers, proposed by DESBOROUGH and HARRIS, 1992; the standard deviation of the controlled variable (*Std(y)*); and the mean of the squared variations in the manipulated variable (*Su*).

4. Results

The control of the first bed in the first reactor of the hydrotreating unit was simulated through block diagrams developed in Simulink/ Matlab 7.0. This simulator was composed by 48 differential equations that represented the bed physical-chemical behavior. In order to evaluate the behavior of the controllers, different scenarios were simulated. For the deterministic study, disturbances were introduced in the feed concentration and in the feed flow rate (± 5% from the respective steady-state values). For the stochastic study, a Gaussian white noise disturbance – e(t) – with variance 1 was introduced in the controlled variable.

Table 2: Comparison between the controllers

Controllers		Deterministic Performance								Stochastic Performance		
		Regulatory										
		Concentration				Flow						
		5%		-5%		5%		-5%				
		ITAE	Su	ITAE	Su	ITAE	Su	ITAE	Su	ANORM	std(y)	Su
GPC	Case 1	11,67	2,09	12,76	1,57	12,72	1,56	11,81	2,60	-	1,13	1,53
	Case 2	10,62	2,74	9,37	1,98	8,94	1,96	11,24	3,42	-	1,56	1,00
	Case 3	8,85	3,82	8,13	2,76	7,80	2,33	9,35	4,77	-	1,00	1,69
ZN	P	114,64	1,00	113,39	1,00	115,19	1,00	112,19	1,00	0,1	1,13	23,61
CC	P	101,13	1,59	100,50	1,63	101,98	1,62	99,08	1,54	0,4	1,15	38,62
	PID	1,00	3,45	1,00	3,41	1,00	3,47	1,00	3,75	0,6	1,37	3623,09
IMC	PI	2,11	1,29	1,91	1,08	1,72	1,06	2,34	1,52	0,3	1,42	15,10
	PID	1,15	2,10	1,16	1,94	1,16	2,03	1,15	2,42	0,4	1,05	938,71
EI	PI	1,65	1,75	1,62	1,55	1,62	1,52	1,73	1,97	0,3	1,44	17,71
	PID	3,48	6,43	3,14	5,79	2,15	5,29	5,24	8,23	0,6	1,76	3697,41

In Table 2, indexes normalized by the smaller value of the column (with exception of the ANORM) can be observed. So, the index equal to 1 in every column represents the smaller value. The cells of the table were colored based on the difference between the respective value and the smallest value of the column: the cells which present values equal to one are green; the cells which present values between one and eight are yellow; and, the cells which present values larger than eight are orange.

The controllers PI and PID – ZN presented unstable behavior, thus they were discarded from this work. The ratio of the pseudo dead time to the process time constant is often referred to as the "uncontrollability" factor (Fc) that is an indication of the quality of control that can be expected (CAMPOS and TEXEIRA, 2006). It is important to realize that the uncontrollability factor of the process is 0.9 and the Ziegler & Nichols method was developed for uncontrollability factor between 0.3 and 0.5 (CAMPOS and

TEXEIRA, 2006). Despite of the fact that the Cohen & Coon method was developed for uncontrollability factors larger than 0.3 (CAMPOS and TEXEIRA, 2006), the controller PI – CC also presented unstable behavior and it was also discarded.

5. Conclusions

The results of this work may also be useful for other chemical engineering applications as the simulated process was identified using a first order plus time delay model, which is a rather typical model in this area. In terms of deterministic performance, some PIs and PIDs presented unstable behavior. The stable ones achieved smaller ITAEs than the ones obtained using GPC. The Case 3 GPC presented an average ITAE nine times larger than the best obtained ITAE. Nevertheless, the Case 3 GPC reduced the controlled variable variability in 41%, requiring an effort in the control action (*Su*) 88% smaller than the smaller effort required among the feedback controllers (IMC-PI), which achieved a reduction of 16% in this variability. The largest achieved reduction in the controlled variable variability among the feedback controllers was 38% (IMC-PID) which required control actions 550 times more intense than the Case 3 GPC. Considering the ANORM index, the feedback controller that achieved a stochastic performance closer of the Minimum Variance Control (MVC) was ZN-P. CARELLI (2008) presents the figures of all the simulations performed. These figures are not reproduced here for space reasons, however Table 2 sums up the main results of the simulations. The performance evaluation for the SISO case studied indicates the GPC controller as the one that provides the best balance, taking into consideration deterministic and stochastic performances, as it is more equipped to deal with random disturbances (due to the CARIMA model) at the same time that provides an adequate deterministic behavior. The superiority of the model predictive controller can also be noticed by the facts that some of the feedback controllers rendered unstable behavior and that the focus of this work was not on finding optimum tuning parameters for the GPC controllers, but parameters that resulted in stable behavior and good regulatory behavior, both in the presence of deterministic and stochastic disturbances.

References

A. C. CARELLI, 2008, Stochastic and deterministic performance assessment of PID and MPC controllers: application to a hydrotreater's reactor – Programa em Tecnologia de Processos Químicos e Bioquímicos, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 111 p.

D. W. CLARKE, C. MOHTADI, P. S. TUFFS, 1987, Generalized predicitive control – part I: the basic algorithm. Automatica, v.23, n. 2, p. 137 – 148, ago.

D. W. CLARKE, 1988, Application of generalized predictive control to industrial processes, IEEE Control Systems Magazine, p. 49 – 55, abr.

H. P. CARNEIRO, 1992, Robust control of a fixed bed chemical reactor, Dissertation of Master Title in Chemical Engineering – Coordenação dos Programas de Pós-Graduação de Engenharia (COPPE), Universidade Federal do Rio de Janeiro, Rio de Janeiro, 142 p.

L. DESBOROUGH, T. J. HARRIS, 1992, Performance assessment measures for univariate feedback control, The Canadian Journal of Chemical Engineering, v. 70, p. 1186 – 1196, dez.

M. C. M. M. CAMPOS, H. C. G. TEIXEIRA, 2006, Controles típicos de equipamentos e processos industriais. Brasil: Edgard Blücher. 396 p. (In portuguese).

V. HLAVÁCEK, 1982, Fixed bed reactors, flow and chemical reaction, Alemanha: Verlag Chemic, p. 103 – 111.

W. MITCHELL, D. SHOOK, 2005, Industrial application of multivariable controller analysis and monitoring techniques, Canadá: Matrikon Inc., 23 p.

1641

Model-Based Fault Diagnosis Using a Hybrid Dynamic Simulator: Application to a Chemical Process

Nelly OLIVIER-MAGET[a,c], Gilles HETREUX[a,c], Jean-Marc LE LANN[a,c], Marie-Véronique LE LANN[b,c]

[a]*LGC-CNRS, Université de Toulouse, INPT-ENSIACET, 118, route de Narbonne F-31077 Toulouse Cedex 04, France*
[b]*LAAS-CNRS, Université de Toulouse, 7 avenue du Colonel Roche, F-31077 Toulouse, France*
[c]*Université de Toulouse ; INPT, INSA*

Abstract

This work presents a fault detection and isolation methodology for the monitoring of Hybrid Dynamic Systems. This methodology rests on a mixed approach, which combines a model-based method for the fault detection and an approach based on data (pattern matching) for the identification of fault(s). This methodology is integrated within the simulation platform *PrODHyS*, through the development of the module *PrODHySAEM*. The goal of this paper is to underline the potentialities of our approach for the diagnosis of the system. This methodology is illustrated by the studies of diagnosis problems in the field of Chemical Process System Engineering.

Keywords: Fault detection and diagnosis, Hybrid Dynamic Systems, generation of non binary signatures, Manhattan distance, extended Kalman filter, object differential Petri nets.

1. Introduction

Thanks to their large application field, numerous works on Hybrid Dynamic Systems (HDS) deal with modelling, stability and control (Zaytoon, 2001). Among them, some research works focus on the monitoring of these systems and many methods have been developed for Fault Detection and Isolation (FDI). The literature quotes as many fault detection and diagnosis methods as many domains of application (Venkatasubramanian, et al., 2003) and the techniques are generally classified as: methods without models – such as quantitative process history based methods or qualitative process history based methods– and model-based methods which are composed of quantitative and qualitative model-based methods. In our case, a model-based approach has been developed. It exploits the extended Kalman Filter to a hybrid dynamic system. The main idea is to reconstruct the outputs of the system from the measurement using observers or Kalman filters and using the residuals for fault detection (Mehra and Peschon, 1971; Welch and Bishop, 1995; Simani and Fantuzzi, 2006). The purpose is to detect the presence of a fault and to locate the occurrence time. The estimations are compared to the normal variable values and so, deviations are interpreted as faults. Next, the problem is similar to a pattern recognition problem. This paper is organized as follows. The first part of this communication presents the proposed model-based methodology. Next, its implementation is underlined and the main fundamental concepts of the simulation library *PrODHyS* are described. This is followed by a presentation of a modelling

within *PrODHyS*. These concepts and so our approach are exploited through the simulation of the monitoring of a didactic example. Finally, section 6 summarizes the contributions and achievements of the paper and some future research works are suggested.

2. *PrODHyS* Environment

The research works performed for several years within the PSE research department (LGC) on process modelling and simulation have led to the development of PrODHyS (Jourda *et al.*, 1996; Sargousse, 1999; Perret *et al.*, 2004; Olivier-Maget *et al.*, 2008). This environment provides a library of classes dedicated to the dynamic hybrid simulation of processes. Based on object concepts, PrODHyS offers extensible and reusable software components allowing a rigorous and systematic modeling of processes. The primal contribution of these works consisted in determining and designing the foundation buildings classes.
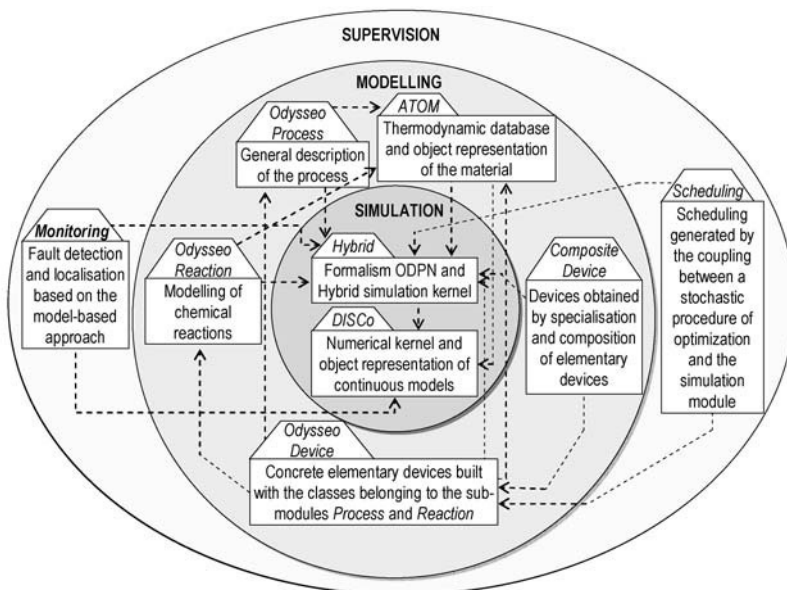
Figure 1. *PrODHyS* Environment

Currently, this library is made up of more than one thousand classes distributed into three independent functional layers (simulation / modelling /supervision) and nine modules (Figure 1). Developed during these research works, the Monitoring module (*PrODHySAEM*) manages the studies of the process monitoring. It is based on our methodology. Thus, each fundamental element of our approach is described by an object class. Moreover the high sequential aspect of the considered systems justifies the use of Petri nets model. This is why the Object Differential Petri Nets (*ODPN*) formalism is used to describe the simulation model associated with each component. A detailed description of this formalism can be found in (Perret *et al.*, 2004; Olivier-Maget, 2007).

3. Supervision module

For this purpose, the simulation model of *PrODHyS* is used as a reference model to implement the functions of detection and diagnosis. The global principle of this system is shown in Figure 1, where the sequence of the different operations is underlined.

Moreover, a distinction between the on-line and off-line operations is made. Our approach is composed of three parts: the generation of the residuals, the generation of the signatures and the generation of the fault indicators.
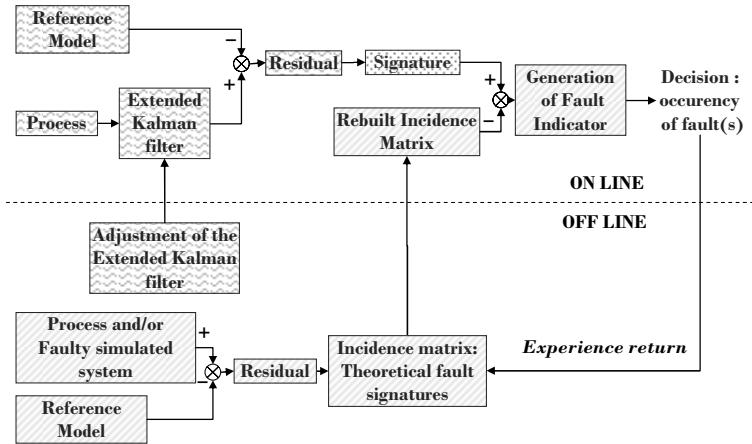
Figure 2. Supervision Architecture

3.1. Generation of the residuals

The first part concerns the generation of the residuals (waved pattern in the Figure 2). In order to obtain an observer of the physical system, a real-time simulation is done in parallel. So, a complete state of the system will be available at any time. Thus, it is based on the comparison between the predicted behavior obtained thanks to the simulation of the reference model (values of state variables) and the real observed behavior (measurements from the process correlated thanks to the Extended Kalman Filter). A description of the extended Kalman filter can be found in (Olivier-Maget *et al.*, 2007). Besides the residual is defined according to the following equation:

$$r_i^r(t) = \frac{\hat{X}_i(t) - X_i(t)}{X_i(t)} \quad with \quad i \in \{1, n\} \tag{Eqn. 1.}$$

where X_i is the state variable, $\hat{X}_i$ is the estimated state variable with the extended Kalman Filter and n is the number of state variables. Note that the generated residual $r_i^r(t)$ is relative. As a matter of fact, this allows the comparison of a residual of a variable with a residual of an other one, since the residual become independent of the physical size of the variable.

3.2. Generation of the signatures

The second part is the generation of the signatures (doted pattern in the Figure 2). This is the detection stage. It determinates the presence or not of a fault. This is made by a simple threshold $\varepsilon_i(t)$. The generated structure $S_i^{rN}(t)$ is denoted by the following equation 2.

$$S_i^{rN}(t) = \frac{Max\left[\left(\left|r_i^r(t)\right| - \varepsilon'_i(t)\right); 0\right]}{\sum_{k=1}^{n} Max\left[\left(\left|r_k^r(t)\right| - \varepsilon'_k(t)\right); 0\right]} \quad with \quad i \in \{1, n\} \tag{Eqn. 2.}$$

with $\varepsilon'_i(t) = \dfrac{\varepsilon_i(t)}{X_i(t)}$, where ε_i is the detection threshold. The value of ε_i is chosen

according to the model error covariance matrix of the Extended Kalman Filter.

3.3. Generation of the fault indicators

The last part deals with the diagnosis of the fault (hatched pattern in the Figure 2). The signature obtained in the previous part is compared with the theoretical fault signatures by means of distance. A theoretical signature $T_{\bullet j}$ of a particular fault j is obtained by experience or in our case, by simulations of the process with different occurency dates of this fault. Then, a fault indicator is generated. For this, we define two distances: the relative Manhattan distance and the improved Manhattan distance. The first distance is denoted by the following expression 3 and the second distance, which allows the diagnosis of many simultaneous faults, is denoted by the expression 4:

$$D_j^{Mr}(t) = \frac{\sum_{i=1}^{n}\left|S_i^{rN}(t)-T_{ij}\right|}{n} \quad \text{(Eqn. 3.)} \qquad D_j^{Ma}(t) = \frac{\sum_{i=1}^{n}\left|S_i^{rN}(t)\times m' - T_{ij}\times n'\right|\cdot T_{ij}}{n'} \quad \text{(Eqn. 4.)}$$

where n' is the number of non-zero elements of the theoretical fault signature $T_{\bullet j}$ and m' is the number of non-zero elements of the fault signature $S^{rN}(t)$.

4. Application
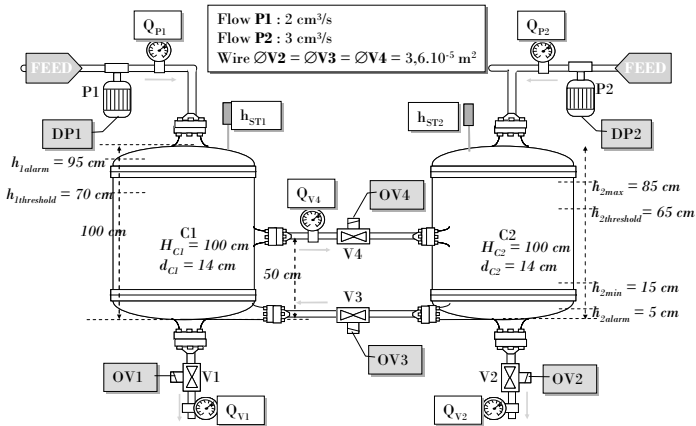
Figure 3. Flowsheet of the benchmark

The considered hydraulic system (Figure 3) is inspired by a benchmark defined by the *AS193* "Diagnosis of the hybrid systems" (cf. www.univ-lille1.fr/lail/AS193/).This system consists of two cylindrical tanks **C1** and **C2**, connected by two pipes with "on/off" valves **V3** and **V4**. The feed of the tanks is maintained by the "on/off" pumps **P1** and **P2**. The tank **C2** can be drained through the "on/off" valve **V2**. The instrumentation of the process is composed of 5 flow sensors and 2 level sensors. The goal of the control device consists in maintaining the liquid level h_2 in **C2** between the heights h_{2min} and h_{2max} by controlling the valve **V4**. The valve **V3** is opened only when the level in **C2** is such $h_2 \leq h_{2alarm}$. The Petri net associated with the control level is presented on figure 4.

4.1. Results

The monitoring of this process is simulated thanks to *PrODHySAEM*. This process is a system based on hydraulic phenomena.
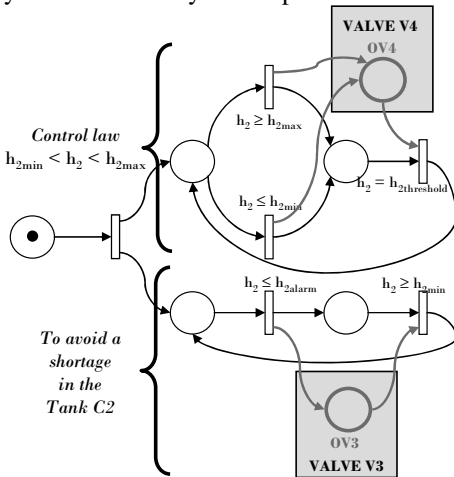
Figure 4. Control Petri net

Two simultaneous faults are introduced: a sensor fault (level sensor of the tank C1, fault 1) and an actuator fault (valve V1, fault 2).

4.1.1. Detection results

We remind that the thresholds for the detection correspond to the model uncertainties obtained by the adjustment of the Extended Kalman filter. Both faults are introduced at t = 1550 seconds. The valve V1 remains opened (fault 1) and the level sensor of the valve V1 returns none value (fault 2).

Figure 5 shows the detection stage. It illustrates the evolution of the residual linked to the liquid level in the tank C1.

From t = 1570 seconds, the residual value of underlines the abnormal behaviour of the process. The diagnosis is launched at t = 1580 seconds.
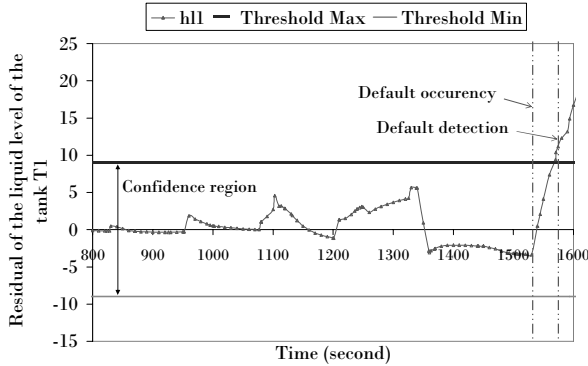
Figure 3. The evolutions of the composition residuals during the evaporation stage

Signature		
	Liquid level h_1	0.01788
	Liquid level h_2	0
	Flow rate in V1	0
	Flow rate in V2	0
	Flow rate in V3	0
	Flow rate in V4	0.98212
	Flow rate in P1	
	Flow rate in P2	0

Table 3. The instantaneous fault signatures

4.1.2. Diagnosis results

The residual is then estimated and we obtain the corresponding instantaneous fault signature (Table 1). Notice that the exploited signature in this approach is non binary, in order to quantify the deviation due to the fault. The construction of the theoretical fault signatures is based on numerous simulations, in which one of the faults exposed in the Table 2 is generated.

We compare the instantaneous fault signature (Table 1) with the theoretical fault signatures, by calculating the relative and improved Manhattan distances (Eqn. 3. and 4.). Then, the fault indicators are generated (Table 2). They correspond to the complement to 1 of these distances.

Fault	Description	Manhattan relative indicator	Manhattan improved indicator
1	**The level sensor of the tank C1 returns none value.**	**0.99676**	**0.99676**
2	**The valve V1 remains opened.**	**0.99954**	**0.99954**
3	The valve V1 remains closed.	0.76170	0.76190
4	The valve V4 remains opened.	0.75447	0.75551
5	The valve V4 remains closed.	0.86747	0.86958
6	The valve V2 remains opened.	0.81459	0.81575
7	The valve V2 remains closed.	0.78329	0.78389
8	The pump P1 provides continuously material.	0.99553	0.87947
9	The pump P1 no more provides any material.	0.77239	0.77279
10	The pump P2 provides continuously material.	0.77444	0.71020
11	The level sensor of the tank C2 returns none value.	0.75447	0.87716

Table 3. The fault indicators of the example

The relative Manhattan indicator points out three faults: the faults 1, 2 and 8 whose indicators are higher than 99%. Nevertheless, any fault is discriminated, since their indicators are higher than 0.68. 0.68 is the fixed criterion, which corresponds to the probability at the standard deviation according to the normal distribution. In the opposite, the results of the improved Manhattan indicator underline the faults 1 and 2 whose are still higher than 99%. Nevertheless, we can not be sure of the presence of either or the both faults. Notice that the simulation of these three scenarios removes the ambiguity. So, this simulation underlines the ability of our monitoring system to detect the simultaneous faults.

5. Conclusion

In this work, the feasibility of using the simulation as a tool for fault detection and diagnosis is demonstrated. Our methodology is based on the hybrid dynamic simulator *PrODHyS*. The fault detection and diagnosis approach, developed here, is a general method for the detection and isolation of the occurency of a fault. Besides, this approach allows the detection of numerous types of fault and has the ability to underline the simultaneous occurency of many faults. The works in progress aim at integrating this simulation model within a model-based supervision system. The goal is to define a recovery solution following the diagnosis of a fault. For this, we exploit the results of signatures in order to generate qualitative information. For example, with these results, we have the ability to distinguish a simple degradation and a failure. Next, we combine our diagnosis approach with an other method, such as classification or case-based reasoning.

References

L. Jourda, X. Joulia and B. Koehret, 1996, Comp & Chem Engineering, 20A, pp. S157-S164
R.K. Mehra and J. Peschon, 1971, Automatica, 5, pp. 637-640
N. Olivier-Maget, 2007, PhD Thesis of the Toulouse University (INSA)
N. Olivier-Maget, G. Hétreux, J.M. Le Lann and M.V. Le Lann, 2008, Chem Eng and Proc, 47(11), pp.1942-1952
J. Perret, G. Hétreux and J.M. LeLann, 2004, Cont Engineering Practice, 12(10), pp. 1211-1223
A. Sargousse, 1999, PhD Thesis of INPT, France
S. Simani and C. Fantuzzi, 2006, Mechatronics, 16, pp. 341-363
V. Venkatasubramanian, R. Rengaswamy, K. Yin and S. N. Kavuri, 2003, Computers & Chemical Engineering, 27, pp. 293-346
G. Welch and G. Bishop, 1995, Technical Report TR 95-041, University of North Carolina
J. Zaytoon, 2001, Systèmes dynamiques hybrides, Hermès Sciences publications

Optimal Control of Heat Exchanger Networks

Luís F. Novazzi,[a] Roger J. Zemp[b]

[a]*Department of Chemical Engineering, Centro Universitário da FEI, Av. Humberto Castelo Branco 3972, São Bernardo do Campo CEP 09850-901, Brazil*
[b]*Faculty of Chemical Engineering, Universidade Estadual de Campinas, Av. Albert Einstein 500, Campinas CEP 13083-852, Brazil*

Abstract

Energy integration among chemical process streams can lead to quite complex heat exchanger networks (HEN) with difficulties in terms of control. In the HEN, the control system primary objective is to keep outlet stream temperatures in a specified range. This objective can be achieved by manipulating bypasses in the exchangers or heat loads in coolers or heaters, which results in a positive degree of freedom, since the number of possible manipulated variables is greater than the number of controlled variables. Thus, a secondary control objective can be set: minimization of utility consumption. In this work HEN dynamics and steady state aspects were addressed with the aim of proposing a control strategy that minimizes utility consumption and satisfies imposed restrictions. With the tools developed in this work it was proposed a feedforward optimal control strategy, which consisted in the minimization of a steady state objective function, connected to utility consumption and subjected to constraints in outlet stream temperatures. The suggested approach is an interesting one since rigid or flexible control objectives can be set. The solution of the minimization problem resulted in optimal bypasses positions in the HEN, which were dynamically implemented using a ramp function and a step function. It was verified that ramp implementation was better, with tolerable dynamic violations.

Keywords: Heat Exchanger Networks, Optimal Control, Dynamics

1. Introduction

Due to the continuous oil price rise since the seventies and to environmental issues, efficient use of energy in chemical processes is very important. Almost thirty years ago the theoretical foundations of Process Integration for the efficient use of energy were established by Linhoff et al. (1982), when they proposed Pinch Technology, an elegant approach to set energy / cost targets for HEN as well as rules to design such networks. Nowadays these synthesis techniques, including some which are based on mathematical programming, are well established in Process Design and are easily found in many Chemical Engineering textbooks.

The design of a HEN depends on nominal streams supply temperatures and flowrates. However, during plant operation such nominal conditions can change, influencing stream target temperatures and propagating in the network. With the aim of studying such propagation, a dynamic mathematical model of the HEN is required. Some dynamic HEN models were suggested in the literature (Mathisen, 1994) and they were based on the resolution of a partial differential equation system.

Since ideally the disturbances must be restrained within the HEN, its control system has as a primary objective to drive stream target temperatures to the specified nominal

conditions. In this regard, it is possible to manipulate bypasses positions in the heat exchangers or to manipulate heat loads in heaters / coolers. There is a positive degree of freedom in doing so since the number of manipulated variables is greater than the number of controlled variables. Due to this fact, a secondary control objective can be set: optimization of utility consumption.

In the paper presented by Boyaci et al. (1996) a control strategy based on the optimization of a steady state objective function satisfying the primary control objective was suggested, but the secondary control objective was not addressed. Glemmestad et al. (1999) presented an alternative approach to the optimal operation of HEN systems based on on-line optimization of a steady state function and a fixed control structure, chosen offline. Lately, Giovanini and Marchetti (2003) have shown that low-level Distributed Control System is also capable of handling HEN control problems when a flexible control loop structure is provided.

Model predictive control (MPC) is an advanced method of process control based on a dynamic model of the plant. In this approach the future moves of manipulated variables depend on the model and on plant measured output variables, in such a way that an on-line constrained optimization is performed. The control problem in the HEN was also considered in the more recent work of González et al. (2006), where MPC in connection to optimization of a steady state HEN model were used, in a two level control structure.

As a simpler way to control HEN when compared to MPC methods, in the present work an optimal feedforward control strategy is proposed, where the HEN utility consumption is minimized, satisfying the secondary control objective, and the primary control objective is written as a restriction in the optimization formulation.

2. Optimal Control in HEN

Quite as important as efficient use of energy in a HEN is the accomplishment of the primary objective control, i.e., stream target temperatures must be in the specified range. In a HEN the variables can be classified as follows: stream supply temperatures T_i^s and stream flowrates m_i are disturbances; bypasses positions f_j and heat loads in coolers or heaters are manipulated variables and stream target temperatures T_i^t are controlled variables. As a consequence of the positive degree of freedom in the HEN, there is a set of different bypasses positions and of heat loads that satisfies the specified target temperatures. Thus, optimal control can be applied to the problem.

2.1. Optimal Control Strategy satisfying the Primary Control Objective
Boyaci et al. (1996) proposed an optimal control strategy, which consisted in the minimization of a steady state objective function, defined by $\sum \left(\delta T_i^t \right)^2$, where the operator δ indicates the difference between a variable and its nominal value. This optimization problem can be written as:

$$\min_{\mathbf{f}} \left(\delta \mathbf{T}^t \right)' \left(\delta \mathbf{T}^t \right) \text{ subjected to } \mathbf{0} \leq \mathbf{f} \leq \mathbf{1} \tag{1}$$

where $\mathbf{f}$ is a vector containing bypasses positions, $\mathbf{T}^t$ is a vector composed of target temperatures and the operator $'$ indicates transposition of a vector or matrix. The restriction imposed in Equation 1 means that the bypasses could be totally closed ($f_j = 0$), fully open ($f_j = 1$) or could be positioned between these extremes. A schematic representation of the methodology is indicated in Figure 1. According to the scheme, disturbance variables $\mathbf{m}$ (flowrates) and $\mathbf{T}^s$ (supply temperatures) are monitored in discrete time intervals and a steady state optimization is performed.
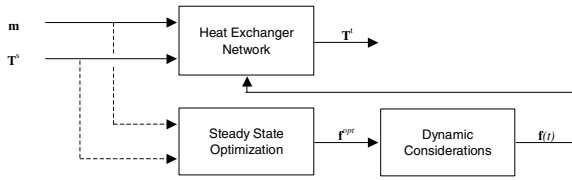
Figure 1 Scheme of the feedforward optimal control strategy

Since the optimization is based on a steady state model, it is necessary to define how bypasses positions change in time. Thus, it was suggested a dynamic implementation (Boyaci et al., 1996) in the form of a step function (Equation 2) and also in the form of a ramp function (Equation 3). In Equations 2 and 3 the bar above variables refers to nominal condition, τ_r is the ramp time duration and t is time.

$$f_i = \begin{cases} \overline{f}_i & \text{for} \quad t < 0 \\ f_i^{opt} & \text{for} \quad t \geq 0 \end{cases} \tag{2}$$

$$f_i = \begin{cases} \overline{f}_i + \left(f_i^{opt} - \overline{f}_i\right)(t/\tau_r) & \text{for} \quad t < \tau_r \\ f_i^{opt} & \text{for} \quad t \geq \tau_r \end{cases} \tag{3}$$

2.2. Proposed Control Strategy
Although the suggested strategy previously described is an interesting one since stream target temperatures are driven to their setpoints, utility minimization is not a concern. In this section an alternative approach is proposed where both primary and secondary control objectives are met.

In a HEN the heaters / coolers are typically positioned in a stream extremity. If δT_i^t assumed a positive value in a hot stream extremity containing a cooler, a higher load in the cooler would be necessary to drive δT_i^t to zero. The same idea applies when a heater is positioned in the extremity of a cold stream. This observation suggests that the optimization control problem can be written to include both primary and secondary control objectives:

$$\min_{\mathbf{f}} \mathbf{p}'\left(\delta\mathbf{T}^t\right) \text{ subjected to } \mathbf{0} \leq \mathbf{f} \leq \mathbf{1} \text{ and } \mathbf{L} \leq \delta\mathbf{T}^t \leq \mathbf{U} \tag{4}$$

where $\mathbf{p}$ is a vector of the same size as $\delta\mathbf{T}^t$ and each $\mathbf{p}$ element is related to a target temperature δT_i^t. The primary control objective, i.e., satisfying temperature setpoints, is represented by $\mathbf{L} \leq \delta\mathbf{T}^t \leq \mathbf{U}$ whereas the minimization function deals with the secondary control objective, i.e., minimizing utility consumption.

Let CP_H and CP_C be the heat capacity for a hot and a cold stream, respectively. The elements of vector $\mathbf{p}$ are filled in as follows: for a cooler in the extremity of a hot stream the element is set to CP_H; for a heater in the extremity of a cold stream the element is set to $-CP_C$ and in the case of absence of a heater or a cooler the element value is set to zero. As a result, the objective function decreases as less hot and cold utility are required. The optimization of the objective function $\mathbf{p}'\left(\delta\mathbf{T}^t\right)$ is subjected to $\mathbf{L} \leq \delta\mathbf{T}^t \leq \mathbf{U}$, where $\mathbf{L}$ and $\mathbf{U}$ are vectors with lower and upper limits for tolerated target temperatures deviations. When there is a rigid control objective, with no tolerable deviations, L_i and U_i are set to zero. On the other hand, when there is a flexible control objective, with relaxed setpoints, L_i and U_i are greater than zero.

It is important to remark that the optimization proposed in Equation 4, which leads to bypasses optimal positions $\mathbf{f}^{opt}$, involves a non linear programming (NLP) problem, with restrictions. The dynamic implementation of $\mathbf{f}^{opt}$ was studied according to Equations 2 and 3, as well as the influence of time τ_r, a very sensitive parameter in the control strategy used in this work.

In terms of implementation, the strategy runs as follows: possible deviations in plant input conditions, like stream supply temperatures and stream flowrates, are monitored and sent to a steady state optimization, based on Equation 4. When this equation is solved and bypasses optimal positions $\mathbf{f}^{opt}$ are found, the signal is sent to the plant and the bypasses will move to their ideal positions. When compared to MPC methods, the proposed strategy herein for HEN control is interesting since it can act as a supervisory system: optimal bypasses positions are fed to local feedback controllers, in a decentralized way.

3. Case Study

With the aim of illustrating the optimal control strategy suggested, it is taken as an example the HEN published by Heggs and Vizcaino (2002), composed by five streams, one heater (H) and four process heat exchangers (HE1, HE2, HE3 and HE4), as represented in Figure 2. In the HEN originally published there was not the presence of bypasses in the exchangers. However, in this work it is supposed that they exist and are positioned in 0,25. As a result, the target temperature vector is equal to: $\mathbf{T}^t = \begin{bmatrix} 111,3 & 133,3 & 177,8 & 175,5 & 204,5 \end{bmatrix}'$. Stream properties and heat exchanger characteristics of this HEN can be found in Heggs and Vizcaino (2002).
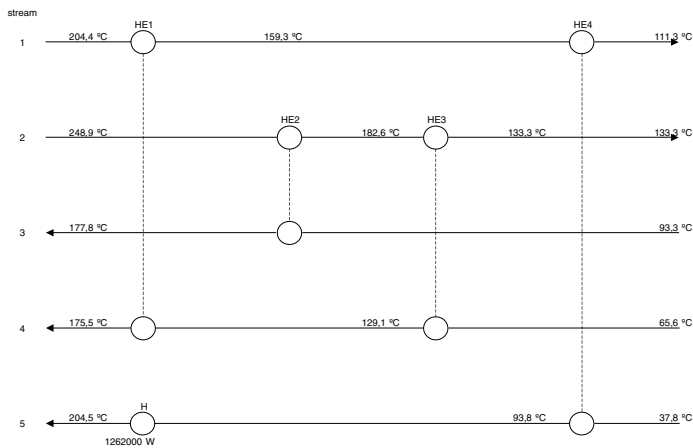
Figure 2 HEN example in grid diagram format

If there is a disturbance in stream 2 supply temperature or flowrate, it will affect the entire network: the heat load in HE2 will change, disturbing stream 3; the heat load in HE3 also will change, interfering in stream 4 and in HE1. Finally the disturbance in HE1 affects HE4, stream 1 and stream 5. To exemplify, let the disturbance be in stream 2 supply temperature, varying from 248,9°C to 258,9°C, i.e., a δT_2^s equal to 10°C, in the instant 200 s. The proposed approach was applied and it was considered a rigid control objective, with $\delta \mathbf{T}^t = \mathbf{0}$. By solving Equation 4, with $\mathbf{p} = \begin{bmatrix} 0 & 0 & 0 & 0 & -11400 \end{bmatrix}'$, one gets the following optimal bypasses positions:

$$\mathbf{f}^{opt} = [0{,}4132 \quad 0{,}2975 \quad 0{,}3540 \quad 0{,}2922 \quad 0{,}2131 \quad 0{,}0427 \quad 0{,}0875 \quad 0{,}0000]'$$

The constrained optimization was performed in Matlab and a steady state model of the HEN was used (Novazzi, 2007). The first and second elements of $\mathbf{f}^{opt}$ represent bypasses positions of hot and cold sides of HE1, respectively, and so on until HE4. The dynamic implementation of $\mathbf{f}^{opt}$ was firstly studied by using a step function (Equation 2) as soon as the disturbance was detected, in 200 s. The dynamic responses are indicated in Figure 3 and they were obtained in Matlab by solving a system of partial differential equations (Novazzi, 2007).
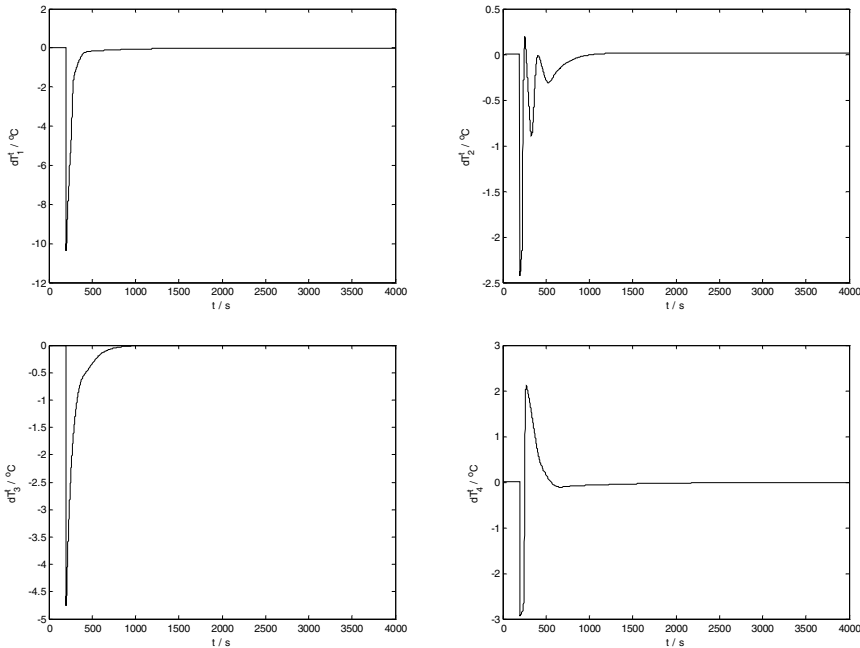
Figure 3 HEN dynamic responses for bypasses opening in a step function

The target temperatures $\delta\mathbf{T}^t$ oscillate before they eventually settle to the former steady state, where $\delta\mathbf{T}^t = \mathbf{0}$, showing that the primary control objective was achieved. Besides, the secondary control objective was also accomplished, since hot utility consumption was reduced from 1,262 kW to 1,096 kW. However, during the transient response severe violations in $\delta\mathbf{T}^t$ happened, mainly in stream 1, with deviations that surpassed 10ºC. Stream 5 was not included in the analysis since it was supposed that any disturbance could be easily met with hot utility in heater H. In spite of the good results in steady state, the responses were very poor dynamically, with severe violations of the restriction $\delta\mathbf{T}^t = \mathbf{0}$.

As a second trial, the bypasses were opened according to a ramp function (Equation 3), with different durations τ_r. The new dynamic responses are indicated in Figure 4, where the solid line represents a τ_r of 500 s, the discontinuous line a τ_r equal to 1000 s and the dashed line a τ_r of 2000 s. The HEN dynamic responses for bypasses opening in a ramp function are clearly superior when compared to bypasses opening in a step function. In stream 1, for example, the dynamic violation of δT_1^t was reduced to only 1,3ºC for a τ_r of 2000 s. It is a significant reduction in the dynamic violation since previously it was greater than 10ºC (Figure 3).
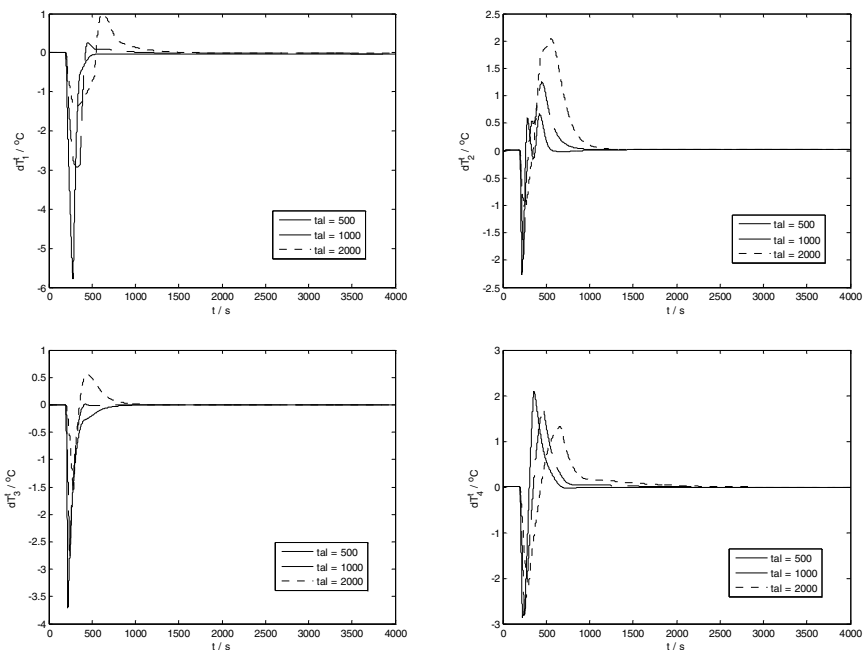
Figure 4 HEN dynamic responses for bypasses opening in a ramp function

4. Conclusions

In this work a feedforward optimal control strategy was suggested and applied in HEN. It consisted in the minimization of utility consumption through bypass manipulation and at the same time stream target temperatures were kept at their setpoints, in a constrained optimization formulation (Equation 4). The proposed approach was successfully applied to a HEN with five streams and the primary and secondary control objectives were met. Dynamic implementation of the optimum bypasses positions showed that step functions led to poor responses, with severe dynamic violations (Figure 3) whereas the bypass opening in a ramp function led to quite satisfactory results (Figure 4).

References

C. Boyaci, D. Uzturk, A.E.S. Konukman, U. Akman, 1996, Dynamics and Optimal Control of Flexible Heat Exchanger Networks, *Computers and Chemical Engineering*, v.20, p.S775-S780

L.L. Giovanini, J.L. Marchetti, 2003, Low-level Flexible Structure Control Applied to Heat Exchanger Networks, *Computers and Chemical Engineering*, v.27, p.1129-1142

B. Glemmestad, S. Skogestad, T. Gundersen, 1999, Optimal Operation of Heat Exchanger Networks, *Computers and Chemical Engineering*, v.23, p.509-522

A.H. González, D. Odloak, J.L. Marchetti, 2006, Predictive Control Applied to Heat Exchanger Networks, *Chemical Engineering and Processing*, v.45, p.661-671

P.J. Heggs, F. Vizcaino, 2002, A Rigorous Model for Evaluation of Disturbance Propagation through Heat Exchanger Networks, *IChemE*, v.80, p.301-308

B. Linnhoff, D.W. Townsend, D. Boland, G.F. Hewitt, B.E.A. Thomas, 1982, *A User Guide on Process Integration for the Efficient Use of Energy*. London: Warwick Printing Company Ltd.

K.W. Mathisen, 1994, *Integrated Design and Control of Heat Exchanger Networks*. University of Trondheim, 195 p., Thesis (D)

L.F. Novazzi, 2007, *Dinâmica e Controle de Redes de Trocadores de Calor*. Universidade Estadual de Campinas, 138 p., Thesis (D)

Soft-Sensor for Real-Time Estimation of Ethanol Concentration in Continuous Flash Fermentation

Elmer Ccopa Rivera, Daniel Ibraim Pires Atala, Aline Carvalho da Costa,

Francisco Maugeri Filho, Rubens Maciel Filho

School of Chemical Engineering, State University of Campinas, P.O. Box 6066, 13081-970, Campinas, SP, Brazil

Abstract

In this work, Artificial Neural Network-based software-sensors (ANN-SS) have been developed for the on-line estimation of ethanol concentration in a continuous flash fermentation process at laboratory scale. The process consists of three interconnected units: fermentor, filter (tangential microfiltration for cell recycling) and vacuum flash vessel (for the continuous separation of ethanol from the broth). The concentration of ethanol in the fermentor and of ethanol condensed from the flash are successfully monitored on-line using ANN-SS calibrated with laboratory experimental data. Results exhibited acceptable predictions for both concentrations of ethanol in the fermentor and condensed ethanol with correlation coefficients of 0.82 and 0.91. The proposed ANN-SS represents an accurate model-based approach which is expected to contribute to improve the implementation of suitable operating strategies of optimization and advanced control to achieve high operation performance.

Keywords: Bioreactor, modeling, ethanol, software sensors, artificial neural network.

1. Introduction

In continuous flash fermentation, the performance of the whole process is significantly influenced by the relationships between process variables in the reaction and separation units. Thus, the availability of an accurate mathematical model is important to improve process performance and to define the most suitable operating conditions to achieve a particular objective. Another important application of an accurate mathematical model is the monitoring of the process states, which is an information source for the decision making in the ethanol production process. In general, monitoring of ethanol profiles is carried out as off-line analysis, often with a significant time delay between sampling and availability of the analysis results.

In industrial fermentation plants analytical instruments are difficult to calibrate, mainly due to the characteristics of industrial culture media, such as turbidity of the culture, presence of dissolved CO_2, among others [1, 2]. Nowadays the software sensor is the preferred alternative for monitoring state variables in biotechnological processes [3].

The so-called software sensors are algorithms for on-line estimation of state variables and model parameters that are not measurable in real-time [4]. This approach is a promising research area with significant impact on industrial operation [5].

More recent works on software sensor for solving biotechnological complex problems are presented by Osorio et al. [2], Lee et al. [6] among others. The major purpose of using software sensors in bioprocesses is to assess the quality of the final product and to

validate on-line analyzers, providing redundant measurements. Artificial Neural Networks (ANN) has been dominant in literature in the field of software sensor design. Artificial Intelligence techniques such as ANNs have been widely applied for bioprocess modeling, monitoring and control. This technique is sought to efficiently combine all available knowledge and to direct the development toward an improved process operation strategy. Besides, ANNs can be used to offer adaptive solutions, since the reestimation of their parameters is a straightforward procedure [7]. These characteristics are suitable for analyzing data from more complex processes such as the continuous flash fermentation, which have a large number of state variables.

In this work a relatively simple ANN-based software-sensors for monitoring a continuous flash fermentation process is proposed. On-line measurements, such as temperature, dilution rate, pH and turbidity in the fermentor, as well as temperature, pressure and feed flow rate in the flash vessel, are considered as inputs to the software sensor used to infer ethanol concentration in the fermentor and ethanol concentration in the condensed stream from the flash vessel. The ANN-SS is evaluated considering the accuracy with which it describes the experimental dataset.

2. Case study: Continuous flash fermentation

2.1. Process Description

Figure 1 show a schematic diagram of the continuous flash fermentation process investigated in this study. This process consists of three interconnected units: fermentor, cell recycle system (tangential microfiltration) and vacuum flash vessel (ethanol-fermented broth separation unit).
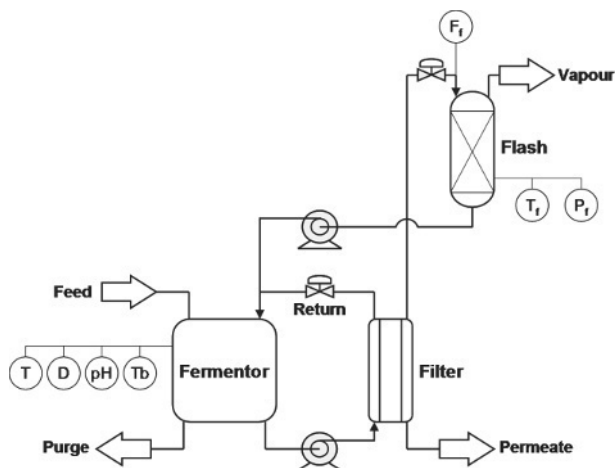
Figure 1. Schematic diagram of the continuous flash fermentation process

The total volume of the system was approximately 5 L. The following on-line variables are taken into consideration in this study: Pressure in the flash, P_f (mmHg); Flash feed flow rate, F_f (L/h); Temperature in the flash, T_f (°C); Temperature in the fermentor, T (°C); Dilution rate in the fermentor, D (h^{-1}); pH in the fermentor, pH; Turbidity in the fermentor, Tb (%).

The process is started-up with the addition of the inoculum in a solution of sugar molasses from an industrial plant diluted to 180 g/L of total reducing sugars. It is operated in semi-batch mode until all the substrate is consumed, with addition of diluted molasses only to complete the volume of ethanol evaporated in the flash vessel. The

objective is that biomass reaches a high concentration before the continuous fermentation begins. The end of the start-up stage is monitored by the stabilization of the turbidity as well as of the condensate volume readings. The continuous extractive fermentation is initiated by turning on a peristaltic pump that removes the permeate in the filtration system. The feed flow rate is adjusted to maintain a constant volume in the system. The dilution rate is adjusted to 0.1 h^{-1}, which correspond to a residence time of 10 hours.

After the steady state is attained, the biomass concentration is 30 g/L (cell viability of 95%) when the feed substrate concentration is of 180 g/L. This steady state is maintained during three times the residence time (in this case, about 30 hours). After the end of this stage, the input and output streams are adjusted to provide a dilution rate of 0.15 h^{-1}, which corresponds to a residence time of 6.67 hours, under the same feed substrate concentration and cell volume in the fermentor. This procedure is repeated for the dilution rates of 0.20 and 0.35 h^{-1}, which correspond to residence times of 5 hours and 2.85, respectively. The flash vessel is operated with a feed flow rate of 200 L/h, pressure ranging from 115 to 205 mmHg and temperature varying from 33.8 to 34.5°C, keeping the ethanol concentration in the fermentor between 50 and 65 g/L. The flash vessel works at this temperature to eliminate the need of heat exchangers (which is the most expensive equipment in alcoholic fermentation plants), since the heat of fermentation is removed from the fermentation broth in the flash vessel during the vaporization of part of the ethanol.

The continuous extractive fermentation has shown several advantages, such as low vinasse generation due to the possibility of feeding molasses at higher concentrations, which reduces costs in waste treatment, and the potentiality of eliminating one distillation column from the process. Further detail of technical features can be found elsewhere [8].

3. Artificial neural network-based software sensor

In the present study, two ANNs, one for each output (concentration of ethanol in the fermentor, P_{ferm}, and condensed ethanol from the flash, P_{flash}). Tests in the current study and previous investigations [7] have shown that this approach lead to better results avoiding unnecessarily complex structures.

When ANNs are used to build inferential measurement systems, a study of their architecture is crucial in order to provide a reduction in the dimension of the input space, which can remarkably reduce the time needed for training [5]. Another important issue is to identify among large numbers of possible input variables the ones that are significant for the desired response.

In this work, a Plackett-Burman (PB) design [9] is used to test several combinations of the input variables and their effects on the ANN-SS prediction accuracy. The set of variables was selected using prior knowledge of the continuous flash fermentation process. An array of primary sensors provides seven on-line measurements (see Figure 1): P_f; F_f; T_f; T; D; pH and Tb. The PB design allows to test up to N-1 factors (design variables) in N trials, where N is a multiple of 4. When the number of examined factors is smaller than N-1, the design is completed with dummy variables, which enables the calculation of the standard error for the factors.

In the PB design used in this study, the input variables were varied at two levels; high (presence) and low (absence). A design with 12 trials was used, which leads to 4 degrees of freedom, considering that there are seven input variables. A total of 12 architectures were tested using the input variables corresponding to the PB matrix.

The appropriate number of neurons in the hidden layer of the 12 architectures was found by the cross-validation technique in order to avoid model over-fitting and to achieve good generalization from the training dataset. This technique splits the data sample into a training dataset and a validation dataset. Then MLP neural networks with different numbers of hidden nodes are trained with the training dataset, and their performance are evaluated on its ability to make correct predictions of the validation dataset in terms of mean square error (MSE) (equation 1). The lowest MSE of the validation dataset were used as response, where the corresponding number of hidden neurons is also reported. The effects were calculated using the software Statistica 7.0 (Statsoft).

$$MSE = \sum (d_k - g_k)^2 \qquad\qquad\qquad\qquad (1)$$

In Equation 3, g_k is the prediction of the neural networks and d_k is the desired output, which in this study are P_{ferm} and P_{flash}.

In this study, both input and output data to the ANN were normalized to the range [0.1, 0.9]. Small random values are used to initialization of weights and biases. Subsequently, the standard backpropagation learning algorithm [10], based on a gradient descendent method implemented in FORTRAN is employed to train each network describing ethanol concentrations. Training was stopped after 2000 epochs.

4. Results and Discussion

Two representative datasets containing 4000 input/output patterns were presented in a randomized sequence to the two neural networks, one for each output (P_{ferm} and P_{flash}). These patterns were randomly selected of a database containing 4420 patterns corresponding to approximately 200 hours of continuous flash fermentation at different dilution rates (0.1, 0.15, 0.20 and 0.35 h^{-1}). The validation, i.e., the predictive capability of the neural networks, was assessed on a different sequence of experimental observations, comprising the remaining 420 patterns of the database.

The development of the ANN-SS started with the evaluation of the influence of the considered input variables on the MSE of the desired responses using PB design. Various combinations of a total of seven input variables were tested to build the software-sensors. It was be seen that the ANN-SS that uses the seven inputs led to the lowest MSE values.

Figure 2 shows the effects of the input variables in the accuracy of prediction of the software sensor for the two responses. For the concentration of ethanol in the fermentor, the highest effect was that of the fermentor temperature, followed by the effects of flash temperature and flash pressure. For the concentration of condensed ethanol, the most important effects were that of flash temperature and dilution rate. These results show the strong influence of input variables from the flash vessel in the fermentor and vice-versa.

The performance of the ANN-SS with the lowest MSEs for the P_{ferm} and P_{flash} are shown in Figures 3a and b, respectively. The dilution rates corresponding to the data are shown in dashed lines. It can be seen that the ANN-SS has a good agreement with the experimental validation data. The measured data versus the prediction values of concentrations of ethanol in the fermentor and condensed ethanol concentration are shown in Figures 4a and b. These plots allow visualizing the prediction quality of the software sensors in a clear way. The correlation coefficients (r^2) values represent the fraction of the variance explained in the experimental observation by the ANN-SS.
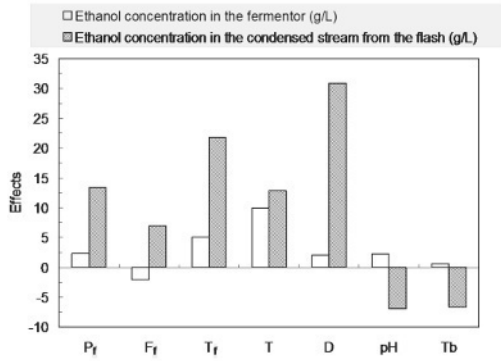
Figure 2. Effect of the input variables on P_{ferm} and P_{flash} from results of Plackett-Burman design
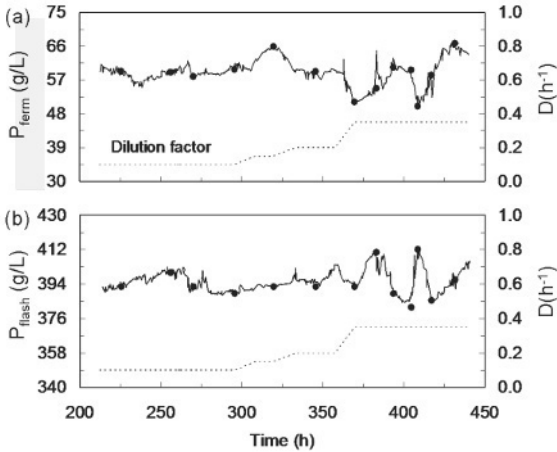
Figure 3. Experimental (solid symbols) and performance of the ANN-Software sensor (continuous lines) using the validation dataset: (a) Ethanol concentration in the fermentor and (b) Ethanol concentration in the condensed stream from the flash. The dilution rates corresponding to the data are shown in dashed lines
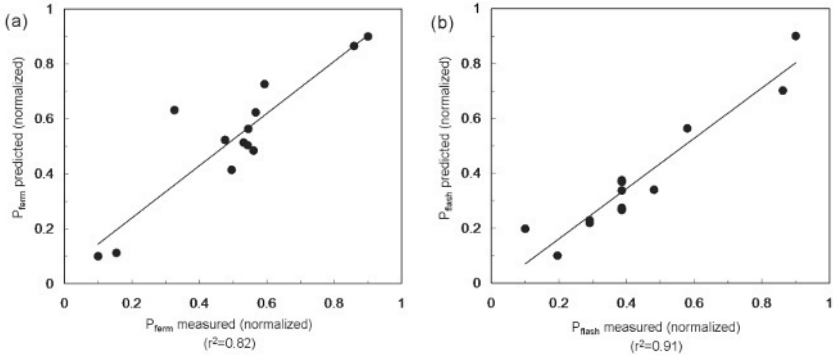
Figure 4. Regression plot from the best ANN-SS architecture during validation: (a) Ethanol concentration in the fermentor and (b) Ethanol concentration in the condensed stream from the flash

5. Concluding Remarks

The development of ANN-based soft-sensors applicable to a continuous flash fermentation process at laboratory scale is addressed in this work. Plackett-Burman design was used to evaluate the influence of various considered input variables in the concentrations of ethanol in the fermentor and of condensed ethanol from the flash vessel. The results have shown that it is possible to accurately infer these concentrations from easily measurable input variables. An important additional advantage of this approach consists in the use of a system of (relatively) cheap sensors such as thermocouples, pH meter and turbidimeter in order to infer the concentrations of ethanol.

The analysis of the effects of the input variables in the concentrations of ethanol in the fermentor and condensed from the flash vessel confirmed close relationships between process variables in the reaction (fermentor) and separation (flash vessel) units. The use of software sensor based on ANNs represents thus an accurate model-based approach which is expected to contribute to improve the implementation of suitable operating strategies of optimization and advanced control to achieve high operation performance.

6. Acknowledgements

References

Macedo, 2003, Estado da arte e tendências das tecnologias para energia. Brasília: Ministério da Ciência e da Tecnologia. CTEnerg Secretaria Técnica do Fundo Setorial de Energia. CGEE Centro de gestão e Estudos estratégicos.

D. Osorio, J.R. Pérez-Correa, E. Agosin, M. Cabrera, 2008, Soft-sensor for on-line estimation of ethanol concentrations in wine stills, J. Food Eng., 87, 571-577.

S.I.T.A. Soons, M. Streefland, G. van Straten, A.J.B. van Boxtel, 2008, Assessment of near infrared and "software sensor" for biomass monitoring and control, Chemom. Intell. Lab. Sys., 94, 166-174.

G. Bastin, D. Dochain, 1990, On-line Estimation and Adaptive Control of Bioreactors, Elsevier, Amsterdam.

J.C.B. Gonzaga, L.A.C. Meleiro, C. Kiang, R. Maciel Filho, 2009, ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process, Comput. Chem. Eng., 33, 43-49.

M.W. Lee, S.H. Hong, H. Choi, J.H. Kim, D.S. Lee, J.M. Park, 2008, Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors, Process Biochem., 43, 1107-1113.

E.C. Rivera, A.C. Costa, R.R. Andrade, D.I.P. Atala, F. Maugeri Filho and R. Maciel Filho, 2007, Development of adaptive modeling techniques to describe the temperature-dependent kinetics of biotechnological processes, Biochem. Eng. J., 36, 157-166.

D.I.P Atala and F. Maugeri Filho, 2004, Processo fermentativo extrativo a vácuo para produção de etanol, C12P 7/14, C12R 1/865. Brazil, PI 0500321-0.

R.L. Plackett, J.P. Burman, 1946, The design of optimum multifactor experiments, Biometrika, 33:305-325.

C.M. Bishop, 1995, Neural networks for pattern recognition, Oxford University Press, Oxford, UK.

1659

Using a Linear APC Controller on a Non-Monotonic Process

Luciana M. Galvão,[a] Leo Lincoln,[b] Chad Segura,[c]

[a] *Luciana M. Galvão, Aspen Technology – Advanced Process Control NALA Professional Services Team, Av. Faria Lima 3729 / 5º, CEP 04538-905, São Paulo - SP, Brazil.*
[b] Leo *Lincoln, Braskem S.A. – UNIB Polo Petroquímico de Camaçari, Rua Eteno s.n., Camaçari – BA, Brazil.*
[c] *Chad Segura, Aspen Technology – Advanced Process Control Global Professional Services, 2500 CityWest Blvd.- CEP 77042- Houston –TX, USA.*

Abstract

This paper demonstrates an application of a linear MPC for the control of a non-monotonic process for which the steady state gain changes the sign depending on the operating conditions. This strategy of using a linear controller on a non-monotonic process is justified if, for instance, there is already a large installed base of linear controllers at the plant site without an existing base of nonlinear controllers and the process belongs to this simple subset of non-monotonic processes. Here, the practical application of this strategy on a cyclopentadiene to di-cyclopentadiene reactor in an isoprene production unit is discussed and the results for this implementation are also presented. The benefit is shown by the reduction in the variability of product and also the reduction of the isoprene losses. In addition, a more stable feed composition to the downstream production unit has been observed. This solution was reached using the Hysys® rigorous steady-state simulator from Aspen Technology, AspenIQ™ for dynamic modeling and the DMCplus® as the linear MPC. Using these tools, it was possible to detect the process non-monotonic behavior and control the dependent variables in a stable manner. This method of using the rigorous model is especially useful when the dead time is big compared to process time constant.

Keywords: non-monotonic process, advanced process control, MPC, cyclopentadiene, Isoprene.

1. Introduction

Nonlinearities are usually faced on Multi-variable Process Control projects. For lower degrees of non-linearity, linear controllers can be adapted to overcome the changes in process gain inherent in nonlinear processes. However, processes which are non-monotonic in nature present a special challenge for APC projects. In a few of these cases, if one can force the process to remain on one side of the zero gain point (point of gain sign inversion), one can usually safely use the linear controller as one has forced the process to remain in a monotonic region. However, if one must operate the process on both sides of the zero gain point such as in processes with uncontrolled disturbances or changes in the controlled variable target, a nonlinear controller is generally advised.

A complete APC (Advanced Process Control) solution has been developed for the Braskem Isoprene Unit. The execution group joins participants from Braskem and Aspentech. The DMCplus® was the primary choice to control all sections of this production site. It would also be required to remain the reaction section with the same control software as the rest of the unit, for keeping better communication between the sub-controllers. The challenge is the reaction section shows a very non-linear behavior in operating window. Due to the process and operating characteristics of the Isoprene reactor, a linear solution for control this system is no longer an option. This issue was solved through the use of simulation (Hysys®), prediction and online (AspenIQ™) tools combined with DMCplus® (MPC Aspen Technology, Inc.). The reaction sub-controller is able to communicate with other DMCplus® sub-controllers and it controls a process where the gains sign change under normal process behavior.

1.1. Process Description

The Braskem Isoprene Separation unit produces Isoprene (99.5%), Dicyclopentediene (DCPD) and Piperylenes. The unit recovers the isoprene (2-methyl-1,3-butadiene) from a feed stream by an extractive distillation process. The feed is C_5 mixed compounds that contents Cyclepentediene (CPD). The product is carried out in a number of process steps. The focus of this work is on the reactor section. The key economic driver for this unit is the isoprene losses reduction.

Since CPD volatility is very close to isoprene's, the separation of these compounds by distillation is practically impossible. By CPD dimerization, it becomes more feasible to purify the isoprene stream and free it of CPD contaminant. The reactor effluent composition has a large effect on the overall process downstream, and affects directly the final product specification. After passing through the reactor, these same reactions keep throughout the unit. Consequently, there is an optimal outlet value different from that of complete conversion of CPD. The optimal value will depend on the overall unit behavior. So, the reactor sub-controller should be able to interact with other unit sub-controllers.

1.2. Reaction

The reaction rates (showed in Figure 1) are affected by feed flow, by the inlet temperature and by CPD concentration in the inlet stream (inlet composition). As the reaction from CPD to DCPD is exothermic, the reactor (plug flow reactor) becomes hotter along the path if the CPD inlet concentration is high. The CPD inlet concentration also affects significantly the reaction rates. If reaction rates are too high, it can reduce isoprene throughput. On the other hand, if there is not enough CPD conversion, then CPD will contaminate the final product (isoprene). The main reaction kinetics model parameters used in the reactor simulation were extracted from literature, and were developed by Seidov (1964) and Muja (1975) at different temperatures ranges.

The reaction rate is function of the concentration and system temperature. It was considered that Arrhenius law can be a good approximation to relate the temperature with the reaction velocity. Because of confidentiality, it was not possible to present the values from F (Frequency Factor) and E (is the activation energy) used in simulation.
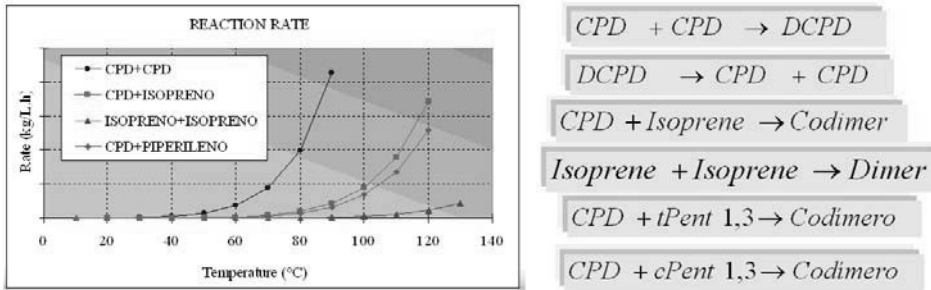
Figure 1: Reactions: homogeneous, in liquid phase, non-catalytic and exothermic. The scale was removed here because of confidential reasons (based on Seidov (1964) and Muja (1975)).

2. Controller Development

2.1. Process Control Challenge and Simulation Development

Figure 2 shows the reactor regulatory control, where the unit's feed stream is heated by a steam heat exchanger. The pressure control will keep the outlet stable and its effect is negligible here, because it is a liquid phase reaction. The only handle for reaction control is the inlet temperature. The inlet CPD concentration is the most relevant disturbance and it affects the temperature along the reactor path. The CPD in the feed changes frequently and abruptly (from 3% to 20%). This disturbance has an inflection point (around 6% CPD), when the steady state gain (disturbance against CPD outlet composition) changes from positive to negative.
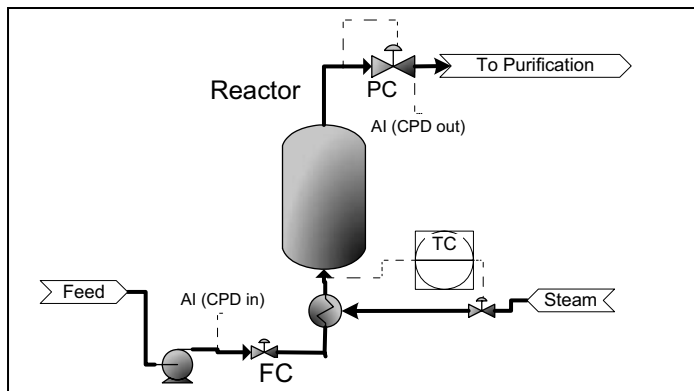
Figure 2 Reactor Regulatory Control,

Figure 3 came from the Hysys® simulation that was developed for this system. The steady-state simulation considered a Plug Flow Reactor, with no catalyst, without axial variation, adiabatic and with exothermic reaction. A number of case studies provided the data for regression. It became finally the online implementation for the control of the CPD outlet concentration. The simulations show that the steady state gain for

temperature (a controller manipulated variable) reverses sign, depending on feed composition (feedforward variable); and also the inflection point in the temperature (where the gain is zero) changes with the feed. It should also be noted that feed rate has an effect on this behavior.

The brown color zone, in Figure 3, should be avoided in process operation because of very high isoprene losses. The green and blue areas are the average conditions where operators try to keep the plant. The blue area is where the isoprene losses are at a minimum and the CPD contamination in the isoprene product gets most close to its limit. In the red area the isoprene will be out of specification. The blue area boundary (the optimum region) depends on downstream capacity and equipment efficiency, so the reactor controller should be able to act properly under various downstream limitations.
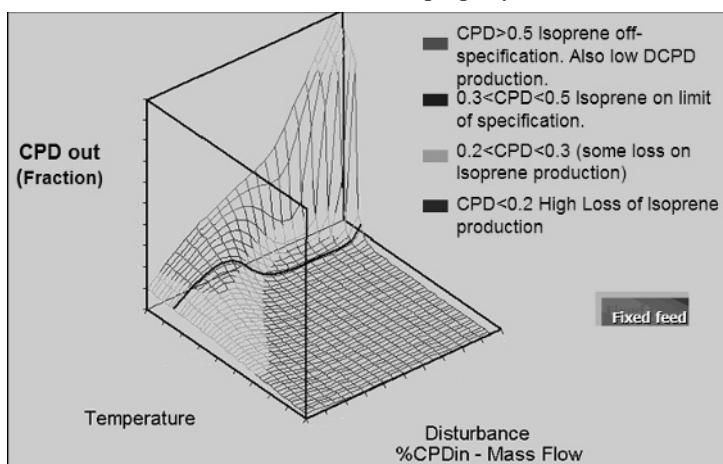
Figure 3. Effect of inlet CPD and inlet temperature on reaction, for various compositions, with fixed flow. Surface obtained by Hysys® Simulation steady state case study.

When CPD inlet concentration is low (below about 6%, for instance), the inlet temperature (control handle) should increase to lower the CPD outlet concentration (increase reaction rate) if CPD in feed rises. However, for high CPD inlet concentration (above 6%, for instance), the opposite action will be required. In other words, with high inlet CPD concentrations, there will be a very high conversion rate (and Isoprene losses as well, because of high reactor temperatures due to the exothermic reaction). In order to compensate for the high feed CPD concentration effect when CPD in the feed is high, temperature control should lower when CPD concentration in the feed rises.

The reactor has a long time to steady state (around 10-15 hours depending on feed rate). In addition, the dead time is approximately 6 hours. A traditional MPC controller with feedback would be virtually unable to control the outlet CPD at a target. The continuous varying controller gains, dead time and sign of the model gains would play havoc with the process.

2.2. Advanced Control Strategy
Figure 4 explains how the simulation was built and how the kinetic model was chosen. In the sequence, the results were checked against real data (at steady state condition)

and then several "Hysys®" case studies" generated the data collection. Finally, a regression was done. The simulation data collection was used to build one temperature equation (by polynomial regression) as a function of: feed, CPD inlet concentration (%), and target CPD outlet (%). The calculated temperature is the required temperature to achieve the target CPD outlet, which is read at CPDout-ET spot inside the controller. The equation for CPD outlet prediction is the same polynomial equation used for temperature, but fed by the current temperature. It predicts the value that CPD outlet would achieve at steady state, if the current temperature remains the same.
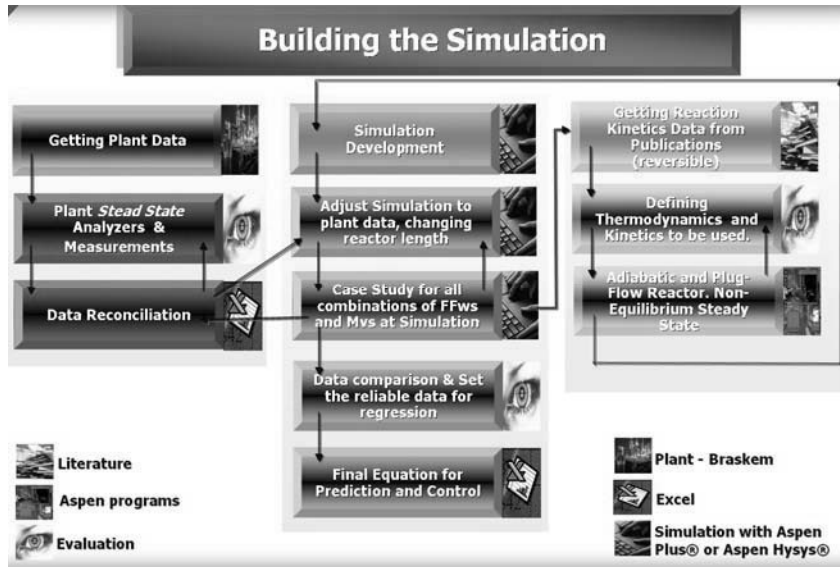
Figure 4: Scheme on how the simulation was developed.

DMCplus® model type is simple model. The controller reads the temperature result, (by IQ™ polynomial calculation) set at the External Target (ET). Then the controller will drive the unit smoothly to that temperature, respecting the other CV limits during dynamics and steady state scenarios. The controller will achieve the temperature-ET and then, at steady state, the CPD current value should be equal to CPDout-ET. The ET is a tool from DMCplus® that corresponds to a set point at the basic control; when driving the unit to the ET, the controller respects the limits and the weighs of other CVs. The CPD outlet ET should be set depending on how the downstream controllers can handle the isoprene losses. In any case, operators are free to specify the CPD outlet ET if required. Inside DMCplus® linear simple model, the CPD outlet variable was configured exclusively to receive the desired value at its ET (by operators or other tool). The CV of CDP outlet is normally OFF and its linear prediction is not used. The appropriate CPD prediction is the polynomial calculation result. The controller will drive to ET temperature, in order to achieve the CPD desired (at steady state).

Figure 5 provides a graphical view of stability rising. The outlet CPD shows a much more stable behavior, following the specified target. The isoprene savings depends on the target value that the operation forces the controller to be.

The model prediction presents a very accurate result when compared to plant data. Most of plant data have the outlet CPD at about 3% or lower. There is not much plant data for

higher values of outlet CPD, (about 5%). The final optimization of isoprene production sometimes requires higher values then 3% (the most usual target). The post controller commissioning stress-test was done. The error (difference between the regression of simulation data and pure simulation curves) average is around 0,14°C. With this new control strategy running, the outlet CPD was safely increased to 4% and some times 4,5%, reducing isoprene losses. Those outlet values were not previously achievable by the plant operators in a stable manner.
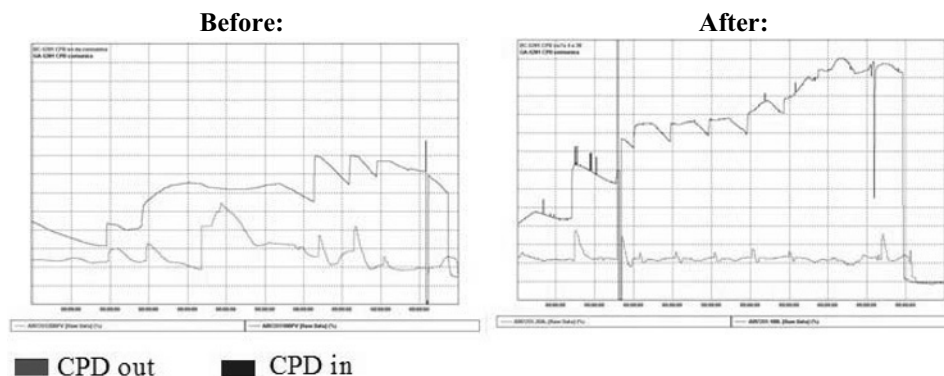
Before: **After:**

Figure 5: The Control Performance after and before the APC implementation. This is a 10 days comparison at same ranges and same scale. The actual numbers on scale were omitted due to confidentiality.

3. Conclusion

This controller is very stable and reaches or surpasses its original expectations. The controller approaches 100% uptime. It also:

- Reduces Isoprene Losses: it was possible to operate the reactor all time in the target temperature to maximize isoprene recovery.
- Demonstrates that rigorous modeling can help to understand the processes and avoid extra-testing in complex systems.
- Shows that if combined with AspenIQ™ and rigorous modeling, DMCplus® can handle non-linear, non-monotonic control for small subsets of variables.

The approach described in this paper can be applied to other processes when there is:
- A long time to steady state process, with a large dead-time;
- Evidence of gains with sign reversal (small controllers only);
- And there exists a simple regressed equation which can adequately predict the process behavior.

4. References

I. Muja, G. Andreescu, M. Corciovei, R. Fratilonium, 1975, Cinetica y Termodinamica unor Reatti Diels-Alder II. Studiul Cinetic al Reatiei de Dimerizare a ciclopentadienei si al reatiilor de codimerizare cyclopentadiena-izopren Revista de Chimie, 26, Nr 12,USA

Seidov N. M., Karirov R. A., Bakhshizade A. A. Azerb Khlm. Zh., 1964, No, 58, 81-86

Bypass Design for Control and Optimization of Heat Exchanger Networks

M. Escobar and J.O. Trierweiler

Group of Intensification, Modelling, Simulation, Control and Optimization of Processes
Department of Chemical Engineering, Federal University of Rio Grande do Sul
Porto Alegre, Brazil (e-mail: escobar@enq.ufrgs.br/ jorge@enq.ufrgs.br)

Abstract

Disturbance propagation during a HEN operation may make the control difficult if the HEN is improperly design. Utility flow rates and bypasses are widely used for effective control of process stream target temperatures, but the number of utility units is usually much less than the number of process streams in the network. This paper addresses the optimal bypass design in heat exchanger networks. It consists in a model-based iterative procedure considering a worst-case disturbance rejection with minimum economic penalty. The methodology is demonstrated using a case study with 3 different structures, making possible a comparison between different options on a quantitative basis, taking into account the optimal operation attainable with minimum total annual cost.

Keywords: heat exchanger network control, controllability, bypass design.

1. Introduction

During the last decades, different approaches were proposed to design the control system in order to accommodate setpoint changes and to reject load disturbances in HENs. Mathisen et al. (1992) provided a heuristic method for bypass placement. Papalexandri and Pistikopoulos (1994 a, b) introduced a systematic framework for the synthesis or retrofit of a flexible and a structurally controllable HEN using a MINLP formulation. Aguilera and Marchetti (1998) developed a procedure for on-line optimization and control system design of a HEN also using a MINLP. Yan et al. (2001) proposed a model-based design for the development a retrofit HEN with optimal bypass placement using a simplified model for disturbance propagation and control.

The procedure proposed in this paper is an evolution of the method introduced by Yan et al. (2001). Our approach considers a more rigorous model for the heat exchangers, an algorithm for automatic pairing selection , and additional bypass are included to reduce the utility consume. Two side effects can occur by the bypass HEN control strategy, since usually increases the (i) interaction and (ii) initial investment. This work analyzes the HEN control, bypass design and minimization of utility consumption simultaneously.

2. Operation and Control of HENs

During HEN operation, degrees of freedom or manipulated inputs are needed for control and optimization. In a HEN with n_s streams and n_u utility units, at least $n_u - n_s$ extra

available manipulations must be used to make the operation structurally feasible, where all target temperatures can be controlled independently. We assume in this work that only a single bypass is used and a stream split is not used as a manipulated variable. In order to deal with positive and negative disturbances, the heat exchanger has to be designed with a steady-state flow rate for the bypass stream different than zero.

For a given HEN, a bypass with a specific nominal value u_{nom} can be added without changing the main HEN structure and operating point if the same heat load needs to be maintained. But besides the opportunity to reject disturbances (*i.e.*, $\delta T^t = 0$), its installation must cause an increment of the heat transfer area (*i.e.*, $\delta A/A_0$). A trade-off between disturbance rejection and costs must be considered during the bypass nominal design. Figure 1 illustrates qualitatively how outlet targets temperatures and the increment of area of the bypassed heat exchanger if the same heat load is maintained.
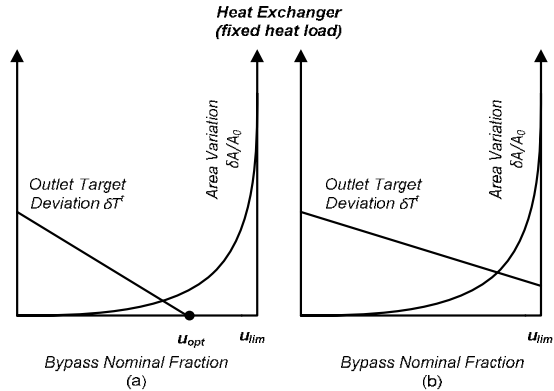
Fig. 1. Qualitative relations between a stream target temperature deviation and the area change as a bypass increases. (a) Complete disturbance rejection when $u_{nom} \geq u_{opt}$; (b) Incomplete disturbance rejection with any bypass fraction.

3. Bypass Design for Control and Optimization

The HEN here is separated into two parts, the inner HEN and the outer HEN according to Glemmestad (1997). While the inner HEN consists of all process exchangers, splitters and mixers, the outer consists of the utility exchangers and target values. The inner HEN contributes with n free variables. For energy optimization the temperatures upstream the utility exchangers, outlets of inner HEN, are to be as close to the stream target temperatures as possible. A model to disturbance propagation and control must characterize the system behaviour under control. It requires a model where the possible control actions are taken into account, represented in matrix form by the equation (1).

$$\delta T^t = G_u \delta u + G_d^t \delta T^s + G_d^w \delta w \qquad (1)$$

The vector δT^t correspond to the deviation target temperatures related to the deviations of supply temperatures (δT^S), heat capacity flowrates (δw), and the bypass fraction by the matrices, G_d^t, G_d^w and G_u respectively. The system model for a given configuration is obtained using a heat exchanger model, structural information relating these units and the steady-state data. The unit model is based in the linearization of the model described by the set of equations (2), and (3) in agreement with the representation in Fig. 2.
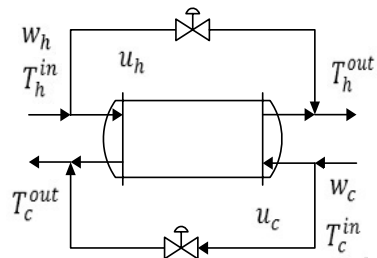
Fig. 2. General structure of a heat exchanger with bypasses.

$$\begin{bmatrix} T_h^{out} \\ T_c^{out} \end{bmatrix} = \begin{bmatrix} \dfrac{R_h - 1}{R_h - a}(1 - u_h) + u_h & \dfrac{1 - a}{R_h - a}(1 - u_h) \\ \dfrac{R_h(1 - a)}{R_h - a}(1 - u_c) & \dfrac{a(R_h - 1)}{R_h - a}(1 - u_c) + u_c \end{bmatrix} \begin{bmatrix} T_h^{in} \\ T_c^{in} \end{bmatrix} \tag{2}$$

$$R_h = \frac{T_c^o - T_c^{in}}{T_h^{in} - T_h^o} = \frac{w_h(1 - u_h)}{w_c(1 - u_c)} \quad and \quad a = exp\left(\frac{UA}{(1 - u_h)w_h}(1 - R_h)\right) \tag{3}$$

For a given HEN, the vectors $\delta T^{s(+)}, \delta T^{s(-)}, \delta w^{(+)}, \delta w^{(-)}, \delta T_{max}^{t(+)}, \delta T_{max}^{t(-)}$ are defined. It is made the assumption of the worst-case design, i.e. the maximum positive and negative deviations of the HEN target temperatures, which occurs at the extreme disturbance values of supply temperatures and heat capacity flow rates if we consider the linear model (eq.1). It is calculated the maximum positive $\delta T_d^{t(+)}$ and negative $\delta T_d^{t(-)}$ deviations of target temperatures considering no control actions (eq. 4 and 5), and it is determined necessary control correction vectors to each case $d^+ = \delta T_{max}^{t(+)} - \delta T_d^{t(+)}$ and $d^- = \delta T_{max}^{t(-)} - \delta T_d^{t(-)}$.

$$\delta T_d^{t(+)} = G_d^{t\,h}\delta T^{s(+)^h} + G_d^{t\,c}\delta T^{s(+)^c} + G_d^{w\,h}\delta w^{(+)^h} - G_d^{w\,c}\delta w^{(-)^c} \tag{4}$$

$$\delta T_d^{t(-)} = G_d^{t\,h}\delta T^{s(-)^h} + G_d^{t\,c}\delta T^{s(-)^c} + G_d^{w\,h}\delta w^{(-)^h} - G_d^{w\,c}\delta w^{(+)^c} \tag{5}$$

It must be selected the subset of at most n manipulated variables δu_s pairing and the subset of n target temperatures that must be controlled using bypass fraction. The pairing is based on the partial RGA ($RGA^\dagger$) defined similarly to the regular RGA substituting the inverse matrix by pseudoinverse of the non square gain matrix G_u (Yang et al., 2001). The basic rule is the same as that in regular RGA, where each controlled variable is paired to a manipulated variable such that the corresponding relative gain is positive and as close to 1 as possible. The best combinations are enumerated.

A pair is selected (δu_s) and the necessary control action $\delta u_s^{(+)}, \delta u_s^{(-)}$, is calculated by solving the optimization problem (min $\|G_u M \delta u_s - d\|_2$) using d^+ and d^- respectively, and the nominal values is updated according to practical limitation $(u_{new} = -min\{\delta u_s^{(+)}, \delta u_s^{(-)}\})$. The convergence criterion is checked $(|\delta u_{new} - \delta u_{old}| \le \varepsilon)$, where ε is the permissible computational error. The model is retrofitted updating the matrices G_u, G_d^t, and G_d^w with the selected nominal bypasses values and all the procedure is repeated until the convergence. At the end the new areas of heat exchangers bypassed are estimated and the model is used to calculate stream outputs deviations and the utility consumption is estimated to the worst case design. The new Total Annual Cost (TAC) is estimated the procedure must be repeated to other possible pair, the best solution is that with the lower TAC.

4. Case Study

In this section 3 different HENs structure depicted in Figure 3, are used to illustrate the design procedure and make a proper selection. The Table 1 lists the design data and the disturbance information need to compute the worst case design.

The degrees of freedom analysis for the $HEN01$, $HEN02$ and $HEN03$ pointed out the number of bypasses that can be used for control and optimization for each HEN ($n = 3$), i.e. the dimension of the space spanned by the bypasses in the target temperature set. A common strategy will use only one bypass to control the temperature target of stream if it is not possible using a utility flowrate. For the $HEN01$ and $HEN02$ only the stream $C1$ must be controlled using a bypass, and two bypasses will be used during operation to minimize the utility consumption. Considering the potential bypass sequence presented in vector $\delta u = \begin{bmatrix} u_{1,1}^h & u_{1,1}^c & u_{1,2}^h & u_{1,2}^c & u_{2,1}^h & u_{2,1}^c & u_{2,1}^h & u_{2,1}^c \end{bmatrix}^T$ the upper limit attainable considering the situation of no driving forces (infinity area) for the $HEN01$ is expressed in the vector $u_{lim} = [0.152 \quad 0.237 \quad 0.495 \quad 0.818 \quad 0.167 \quad 0.083 \quad 0.167 \quad 0.633]^T$ if is assumed a $\Delta T_{min} = 10^0 C$ the vector $u_{lim} = [0 \quad 0.1 \quad 0.43 \quad 0.79 \quad 0.09 \quad 0 \quad 0 \quad 0.56]$. These limits must be considered during the bypass nominal design, associated with a controllability measure to select the appropriated pairing, i.e. taking into account the economic penalty.

With the matrices G_u and $RGA^\dagger$, the potential manipulated variables are sorted by all positive values as close to one as possible creating a priority order. One could think that if we select the prior pair were the best, but this procedure could result in the selection of the same heat exchanger to be bypassed for different controlled outputs, which would not desirable, since there is only one degree of freedom per exchanger. In addition, even if this pair is a possible one, no economical penalty is considered. But the priority order reduces drastically the number of possibilities to be enumerated, and define an appropriated sequence for this enumeration (probably the best pair will be one of the first). The strategy used here consider that keeping the control of the inner HEN This strategy implies that disturbances from the inner to the outer HEN when a utility unit is present are permissible, but to reject disturbances the utility consumption may increase if the disturbance combination results in a near worst case scenario.
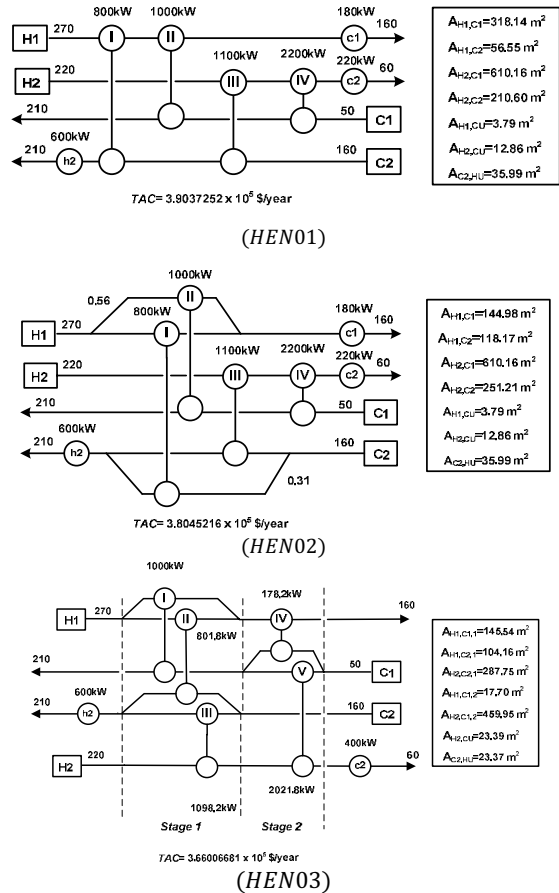
Fig 3. Synthesized HENs for the Case Study using different approaches.

Table 1. Design data for the Case Study.

Stream	T^t (°C)	T^s (°C)	F (kW°C^{-1})	h (kW m^2 °C^{-1})	$\delta T^{(+)}$ (°C)	$\delta T^{(-)}$ (°C)	$\delta w^{(+)}$ (kW°C^{-1})	$\delta w^{(-)}$ (kW°C^{-1})
H1	270	160	18	1	2	-2	3.6	-3.6
H2	220	60	22	1	2	-2	4.4	-4.4
C1	50	210	20	1	2	-2	4	-4
C2	160	210	50	1	2	-2	10	-10
CU	15	20		1				
HU	250	250		1				

Cost of Heat Exchangers ($y^{-1}) $=4000+500[\text{Area (m}^2)]^{0.83}$
Cost of Cooling/Heating Utility $= 20$ ($kW^{-1}y^{-1}) /200 ($kW^{-1}y^{-1})

In order to compare alternatives it is necessary to estimate the range of the utility consumption needed to reject disturbances in the two worst cases. It will provide the range of variation during operation. To avoid this variation, and make a fair comparison between the three alternatives the inner HEN is controlled using as most bypasses as possible according to the rank of the matrix energy balance. The best results obtained to the retrofit of each structure using the bypass design procedure proposed is summarized in the Table 2.

As only 3 bypasses could be manipulated independently, one target temperature is controlled using the utility flow rate, which has been selected based on the cheaper utility. The results show that different investment levels are needed to each case. Comparing the three HENs the last one needs more investment, but the costs associated with the critical consumption assumed by the utility (used to estimate the operating cost) controlled target result in the cheaper solution.
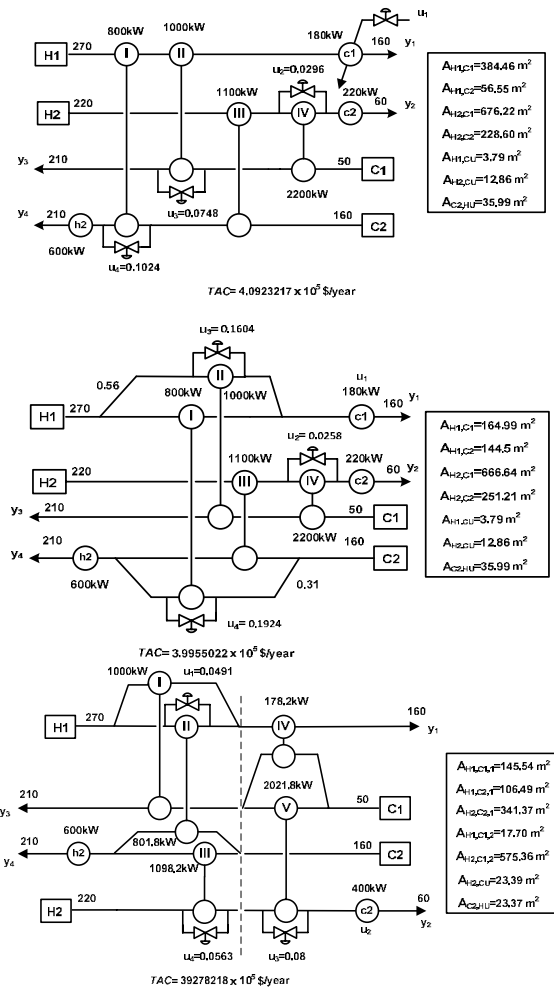
Fig 4. Retofitted HENS using bypass design.

Table 2. Bypass Nominal Design for the 3 structures.

Structure	Controlled set	Manipulated set	Nominal Bypass Value	RGA Number	$(\delta A/A_0)$ %	$I\text{-}I_0$ ($/year)	TAC ($/year)
HEN01	T_{h2}^t	$u_{2,1}^h$	0.0296	0	20.93	22506	409294
	T_{c1}^t	$u_{1,1}^c$	0.0748		10.85		
	T_{c2}^t	$u_{2,2}^c$	0.1024		9.03		
HEN02	T_{h2}^t	$u_{2,1}^h$	0.0258	0	9.29	20572	399550
	T_{c1}^t	$u_{1,1}^h$	0.1604		18.32		
	T_{c2}^t	$u_{1,2}^c$	0.1926		19.23		
HEN03	T_{h1}^t	$u_{1,2,1}^c$	0.0491	1.44	3.92	25811	392782
	T_{c1}^t	$u_{2,1,2}^h$	0.0800		25.27		
	T_{c2}^t	$u_{2,2,1}^h$	0.0563		18.69		

This procedure must reject non controllable HENs, and provide a decision based on the trade-off between controllability and total costs, for this case the cheaper solution shows interaction a little bigger as can be seen by the *RGA* number. The final structures with control schemes are depicted in Fig. 4. For a final decision the dynamic model must be analyzed. The number n is an upper bound on the number of bypasses used to control. The investment cost increases, but it makes possible an operation close to the minimum utility consumption, if the maximum number of bypasses is designed. Therein, there is a trade-off that must be explored.

5. Conclusions

The design of a cost effective HEN capable of be controlled has both economical and operational significance. The controllability depends on the HEN structure, but to be evaluated it is not necessary to design the controller, but it must be selected a set of manipulated and controlled variables. It was presented a systematic model based framework for designing an appropriated control system, selecting the manipulated set, and design bypasses with minimum economic penalty. Through the prediction of the disturbances on controlled variables, it is possible to estimate the bypasses nominal fractions able to reject these disturbances solving an analytical problem per iteration resulting in a fast and robust procedure. The results demonstrate that the 3 structures analyzed are highly controllable with similar total annual cost. In order to make a final decision a comparative controllability analysis including the dynamic performance should be performed.

References

N. Aguilera, and Marchetti, J. L. , 1998, Optimizing and controlling the operation of heat exchanger networks. AIChE Journal, 44(5), 1090–1104.

B. Glemmestad, 1997, Optimal Operation of Integrated Processes, Studies on Heat Recovery Systems. Ph.D. thesis, Norwegian University of Science and Technology, Trondheim, Norwat.

K. W. Mathisen, S. Skogestad, and T. Gundersen, 1992, Optimal Bypass Placement in Heat Exchanger Networks,'' AIChE Meeting, New Orleans, LAŽ1992.

K. P. Papalexandri, and E. N. Pistikopoulos, 1994, A Multiperiod MINLP Model for the Synthesis of Flexible Heat and Mass Exchange Network, Comput. Chem. Eng., 18, 1125Ž.

Q. Z. Yan, Y. H. Yang, and Y. L. Huang, 2001, Cost-Effective Bypass Design of Highly Controllable Heat-Exchanger Networks, AIChE J., 47(10), 2253-2276.

Experimental Study of a Polycondensation Reactor Control by NMPC

Reinaldo A. Teixeira, Galo A. C. Le Roux

LSCP/CESQ - Department of Chemical Engineering, Polytechnic School - University of São Paulo. Av. Prof. Luciano Gualberto, n. 380, trav. 3, CEP 05508-900, São Paulo, Brazil

Abstract

In this work, an implementation of the nonlinear predictive controller (NMPC) to a real 1,25 liter experimental polycondensation reactor is presented. The reaction studied is the reaction of dimethyl terephthalate (DMT) with Ethylene glycol (EG) producing esters and methanol (ME). The objective is to control the reactor temperature by means of the NMPC controller. A simplified model was derived from a complex phenomenological model. The simplified model was adjusted to experimental data. Two different implementations are presented. The performances of the NMPCs are compared with that of the traditional PID controller. NMPC with simple state feedback has the best performance because the state of the process is introduced into the model at each sampling time. Furthermore, NMPC with state feedback had a performance better than the PID in all the trajectories that were studied. The NMPC controller with simple state feedback is flexible because there is no need to adjust the model when there are disturbances in the process. This flexibility is important in the polymer industry, which generally works with multiproduct batch reactors, different batch sizes and variable product qualities.

Keywords: Experimental reactor, nonlinear predictive control, polycondensation, supervisory control.

1. Introduction

The process of batch polycondensation presented in this study is the reaction of dimethyl terephthalate (DMT) and ethylene glycol (EG), generating products such as esters and methanol (ME). The removal of methanol promotes the progress of the reaction to obtain a high degree of conversion. Reactors operating in this regime are known as semi-batch. The operation of reactors in semi-batch mode is a very usual practice in the polymer reaction engineering (Richards and Congalidis, 2006). The major concerns for process control are: safety, product quality, productivity and easy of scaling up (Bonvin, 1998). The supervision and control of the process in batch operation are of great importance in order to achieve these goals, (Ruppen et al., 1997).

Some difficulties encountered in batch processes are the absence of steady-states (natural characteristic of batch reactor) and the uncertainty of the mathematical model (Rho et al, 1998). Polycondensation processes involving the separation of the condensate (which are small molecules produced in the reaction of polycondensation of PET, e.g., water or methanol) have a high degree of nonlinearity when they reach the vapor-liquid equilibrium because of the interactions between the manipulated variables, generating control problems. For this case, the linear predictive controller (LMPC) and the traditional PI and PID may show a poor performance (Pottmann and Seborg, 1996).

Due to the control difficulties in the operation of polycondensation processes, decisions are often left to the operator, which is subject to operational risk (Bonvin, 1998). This justifies the experimental application of NMPC to polycondensation reactors (Henson, 1998, Nagy et al., 2007).

2. Experimental reactor and data acquisition system implementation

The batch reactor has a capacity of 1.25 liters. The system is supervised by a PLC (programmable logic controller). The reactor is jacketed laterally. Air is blown through this jacket in order to cool the reaction system.

The heating is supplied by a resistance at the reactor bottom. There is a thermal resistance (RTDs) for measuring the temperature of this resistance. A second RTD is inserted inside the reactor and measures the temperature of the reaction medium (Tr, reactor temperature). A packed–bed column is attached to the vapor outlet. The top of the column is connected to a partial condenser made up of a jacket through which cooling water flows and inside which vapor condenses in a bundle of vertical tubes. The flow of water used for cooling the partial condenser is handled through a peristaltic pump, also connected to the PLC. Another RTD is inserted into the partial condenser to measure the temperature at which the volatile products are condensed (Tcp, condensate temperature). The process is supervised by the SCADA software IFIX$^\circledR$ 2.20 from GE Fanuc.

The architecture used for supervision and control action calculation consists of a Pentium 1.8 GHz computer with 1GB of RAM connected via RS-232 to the PLC. The IFIX 2.20 program has a VBA (Visual Basic for Applications) integrated, that allows to interact with external programs (eg, an executable NMPC controller). MATLAB 7.2 was used in order to develop the NMPC controller. The NMPC program is an executable file that allows receiving reactor temperature data from the IFIX program and produces an output value that is implemented in the experimental reactor.

The NMPC controller manipulates not only the heating power (Qr), but also the air-cooling flow to the reactor jacket (Qac) and in this case the NMPC has to solve a MINLP (mixed integer nonlinear programming) problem.

2.1. NMPC Algorithm

The NMPC algorithm is based on the minimization of an objective function, J(u), which a quadratic function of the errors of the controlled and manipulated variables that quantifies the performance of the controller within the prediction horizon (P). The control actions vary within a horizon M < P, known as the control horizon (M). Only the first control action computed is implemented in the process. Q, R and S are positive definite weighting matrices that penalize the controlled variables, the manipulated variables and the incremental manipulated variables respectively.

The flow rate of cooling is a binary variable, and the NMPC problem leads to a MINLP. The implementation of NMPC with MINLP problem is solved by exhaustive enumeration. In this case, there are four possible combinations, as the control horizon is 2. The combinations are:

1-heat in all control horizon (two heating actions);

2-heat as the first control action and then cooling;

3-cooling as the first control action and then heat;

4-cooling in the entire control horizon, in this case there is no optimization variable, and then the algorithm only evaluates the value of the function.

The solutions for each of these problems (except the fourth) are calculated using the successive quadratic programming in MATLAB. The lowest value of the four situations corresponds to the control action that is implemented in the process (reactor).

3. Simplified model of the experimental polycondensation reactor

A simplified model was derived from a complex phenomenological model. The complex model in details can be found in Carrillo Le Roux and Teixeira (2004). In the complex model, balances and equilibrium equations are considered for the reactor and the column with the partial condenser at the top.

The simplified model consists of a reactor and one external tank, that simplified represents the column by allowing the reflux of some liquid with a given composition to the reactor. The model assumptions adopted are:
- The column behaves as a tank where Ethylene glycol and methanol are accumulated.
- The hold-up of vapor phase is negligible;
- The heat loss through the walls of the external accumulator is negligible;
- The relation between the molar hold-up and the flow liquid is based on Billet (1995) correlation;
- The composition of the reflux to the reactor is given.

The simplified tank equations are represented by only mass balances.

The energy balance in the reactor, takes into account the thermal inertia of various physical components of the reactor and the resistance, and the heat transfer between these physical components is considered by phenomenological constants. The simplified model was adjusted to experimental data obtained in an open loop test (see figure 1). The test signal was applied in the heating power (Q_r) and cooling flow (Q_{ac}). "Nothing" is applied to regions where no heat or cooling takes place.
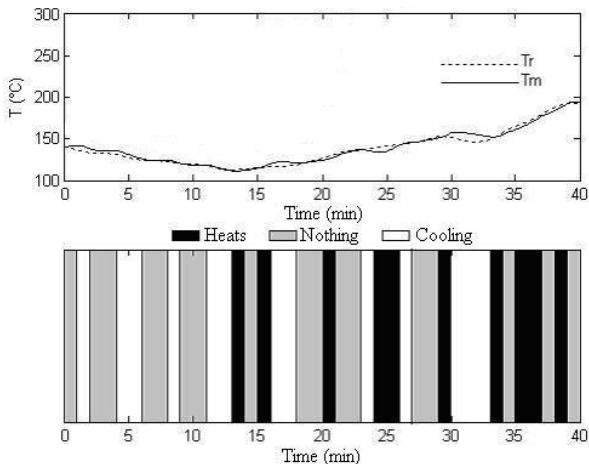
Figure 1 - Simplified Model adjusted with the experimental data obtained in an open loop test.

4. Application of the NMPC

The NMPC controller was used to close the temperature loop of the reactor. For the temperature on the top of the column, a PI controller was used. The controlled variable is the temperature of the experimental reactor (T_r) by manipulating of the heating power

(Qr) and the flow of the cold air flow in the reactor jacket (Qac). To control the partial condenser temperature (Tcp) the cooling flow (Qw) of cold water is manipulated.

The tuning parameters used in NMPC controller are M = 2 and P = 4. These were obtained from simulations where the performance of the controller was compared taking into consideration the necessity of a reduced computational task (Carrillo Le Roux and Teixeira, 2004). The batch duration is 120 min and at the end of the batch the controller is allowed to extrapolate from this point in the calculations. The weighting matrices Q = 1, R = 0, S = 1e-3, the sampling period Δt = 1 minute and the constraints used in inputs are: umin = 0 and umax = 100. For the PI controller, the tuning is: Kp = 0.5%/°C and Ti = 0.01 min/repetitions and sampling period is 1 second. In all the experiments, the reactor is loaded with 1.5 moles of DMT (dimethyl terephthalate) and 3 moles of ethylene glycol (EG). Zinc acetate at a concentration of 3.75e-4 (mol cat/mol DMT) is used as a catalyst. In order to favor the progress of the etherification reaction, the set point for the column condenser temperature is 67 °C, allowing preferably the removal of ME. The trajectory used as the set point for the reactor starts at 140 °C and goes to 180 °C at a rate of 4 °C/min, thereafter the rate is 0.25 °C/min until 200 °C, keeping this temperature until 120 min;

To compare the performance of the NMPC controller a PID was used, with tuning parameters Kp = 1.45%/°C, Ti = 0.28 min/repetition, Td=1.4 min with sampling period 6 s. The comparison is presented in Figure (2) and (3) and the values of the ISE and IAE are presented in Table 1. Comparing the results, the PID controller had a performance equivalent to that of the NMPC controller. Notice that the model (see the model temperature, Tm) in figure (3) and the reactor temperature agree during the steps where NMPC control has a good performance. However, between the times 100 and 120 minutes the model temperature gets far to the reactor temperature (Tr) and the performance of the NMPC controller decreases. To increase the performance of the NMPC controller, a simple strategy of state feedback was implemented: the measured reactor temperature is fed as the initial condition of the reactor temperature in the reference model. The results for the NMPC with simple feedback are presented in figure (4) and table 2. The previous NMPC is thus called simple NMPC. The performance of NMPC with simple state feedback increases significantly and is better than the previous NMPC and PID controllers.
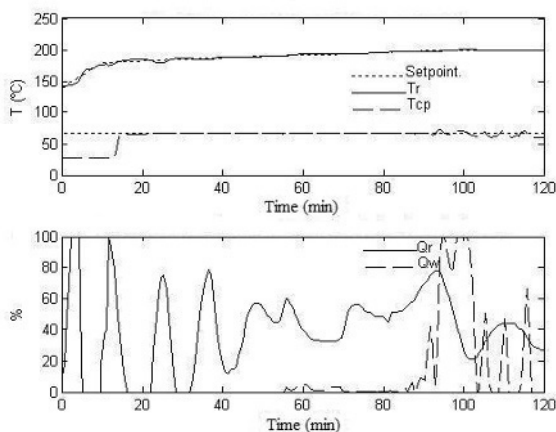
Figure 2 - Results of the PID controller applied to the experimental reactor.
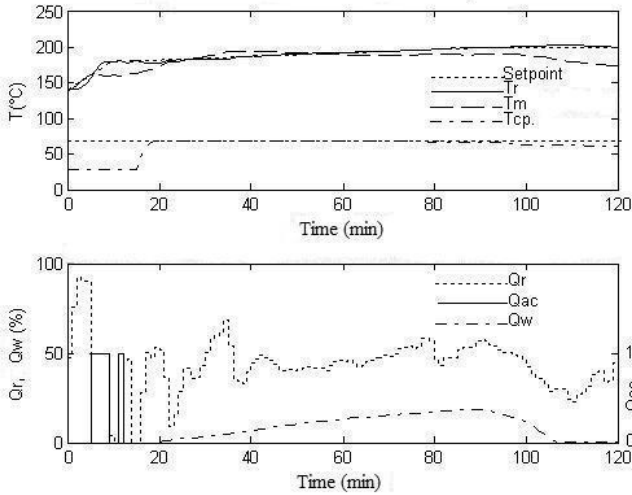
Figure 3 - Results of the simple NMPC controller applied to the experimental reactor.

Table 1 – Comparison of NMPC and PID controllers by ISE and IAE values.

	NMPC	PID controller
ISE	696	621
IAE	198	241
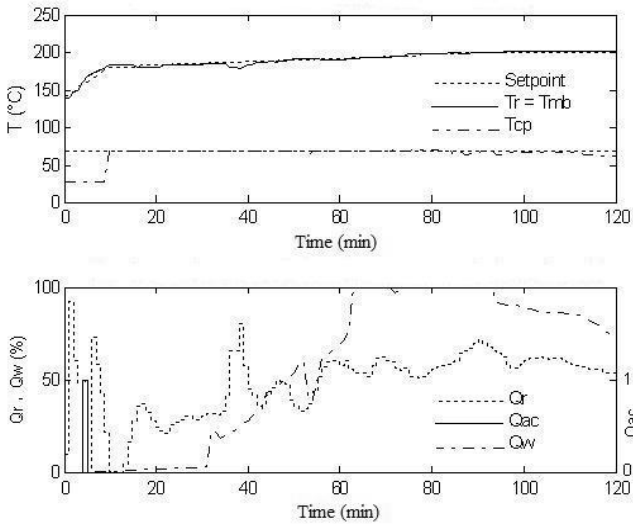
Figure 4 - Results of the NMPC with state feedback in the experimental reactor.

Table 2 – Comparison of the NMPCs and PID controllers by ISE and IAE values.

	NMPC controller with simple state feedback	Simple NMPC	PID controller
ISE	585	696	621
IAE	186	198	241

5. Conclusions

Two NMPC controllers were implemented in an experimental reactor for polycondensation and their performances were compared with that of a PID controller using ISE and IAE metrics. The PID controller showed a performance similar to that of the Simple NMPC controller. The performance of the Simple NMPC controller was degraded because the model does not represent well the process in some ranges. Therefore, an NMPC with simple state feedback was proposed it had a performance better to that of the Simple NMPC and the PID controllers. An advantage of the NMPC with simple state feedback is that it isn't necessary to adjust the model when disturbances in the process take place. This flexibility is important in the polymer industry, which generally works with multiproduct batch reactors, different batch sizes and variable product qualities.

Acknowledgements

References

R Billet,1995, Packed Towers in Processing and Environmental Technology, VCH Publishers, Inc., New York,.

D. Bonvin,1998, Optimal operation of batch reactors - a personal view, J. Proc. Cont., 5-6, p.355-368.

G. A. Carrilo Le Roux; R. A.Teixeira, Application of nonlinear predictive control to a semi-batch polycondensation reactor, Industrial Engineering & Chemistry Research,43(23) p. 7303- 7311, 2004.

M. A. Henson,1998, Nonlinear model predictive control: current status and future directions, Comput. Chem. Engng, 23 p.187-202.

N. Kaistha; M. S.Johnson, C. F.Moore, M. G. Leitnaker. Online batch recipe adjustments for product quality control using empirical models: Application to a nylon-6.6 process, ISA Transactions, Vol. 42 p.305-315, 2003.

Z. K. Nagy, B. Mahn, R. Franke, F. Allgöwer. Evaluation study of an efficient output feedback nonlinear model predictive control for temperature tracking in an industrial batch reactor, Control Engng Pratice,Vol. 15 p.839-850, 2007.

M. Pottmann, D. E. Seborg. A nonlinear predictive control strategy based on radial basis function models, Comput. Chem. Engng., Vol. 21, No. 9, p. 965-980, 1997.

J.R. Richards J. P. Congalidis, Measurement and control of polymerization reactors, Comput. Chem. Engng., Vol. 30 p. 1447-1463, 2006.

D. Ruppen, D. Bonvin, D.W.T. Rippin, Implementation of adaptative optimal operation for a semi-batch reaction system, Comput. Chem. Engng., 22, p.185-199, 1997.

H.-J. Rho, Y.-J. Huh, H.-K. Rhee, Application of adaptive model-predictive control to a batch polymerization reactor, Chem. Eng. Sci., 53, p.3729-3739, 1998.

R. A. Teixeira, Aplicação de controlador preditivo não linear a um reator de policondensação, M.Sc. Thesis (in portuguese), Escola Politécnica da USP, 2003.

Control Loop Performance Assessment and Improvement of an Industrial Hydrogen Generation Unit

Luís G. S. Longhi,[a] Alex S. Reginato,[b] Leandro P. Lusa,[c] Santiago S. Gonzalez,[d] Thiago D. Fleck,[e] Cristhian A. C. Cortez,[f] Herbert C. G. Teixeira[g]

[a,b]*Alberto Pasquilini – REFAP S A., Av .Getúlio Vargas, 11001, Canoas/RS, 92420-221, Brazil, {longhi,alex.reginato}@petrobras.com.br*
[c,d,e]*Trisolutions Soluções em Engenharia, Rua General Bento Martins, 24/111, Porto Alegre/RS, 90010-080, Brazil, {leandro,santiago,thiago}@trisolutions.com.br*
[f,g]*Petróleo Brasileiro S.A. – PETROBRAS, CENPES, Rio de Janeiro/RJ, Brazil, {cristhiancortez,herbertteixeira}@petrobras.com.br*

Abstract

This paper describes the CLPA (Control Loop Performance Assessment) and improvement of an industrial hydrogen generation unit (HGU) located at Alberto Pasqualini – REFAP S.A., a petroleum refinery (whose capacity is 189,000 barrels per day) located at the south of Brazil, subsidiary of Petróleo Brasileiro S.A. (PETROBRAS), and also belonged to REPSOL YPF. At first, we describe the HGU unit and its process purpose. After, we introduce the motivation for the CLPA for this unit and describe the main tool used to carry out this task. This tool is entitled BRPerfX (petroBRas PERFormance indeX) and it was developed by the Federal University of Rio Grande do Sul, Trisolutions enterprise and CENPES (the PETROBRAS research center). The necessary capital to develop this tool was supported by PETROBRAS. A base case was developed for comparison. This base case is composed by a time interval of typical operation before the project startup. This allows showing the improvements made after the control loop assessment is finished. The main steps of the work methodology are also presented, with focus on the planned and implanted actions to improve the dynamical performance of the unit. Finishing the paper, we present the main results achieved. These include not only the better dynamic behavior of the loops but also the economic results related to the index online monitored in the CLPA tool.

Keywords: control loop performance assessment, process control, industrial application.

1. Introduction

Since the publication of the seminal work of Harris [1], the interest in control loop performance assessment (CLPA) has reflourished in the control community. This occurred because, for the first time, this community clearly saw that was possible to deal with the important problem of online evaluating the performance of control loops with a realistic framework. In subsequent years several works concerning this subject arose in the open literature and some tools started to be developed by some control and automation enterprises. Instead of purchasing some of those tools, PETROBRAS, at dawn of current millennium, decided to support Federal University of Rio Grande do Sul and the Trisolutions enterprise aiming the development of a new tool for CLPA.

This new tool has evolved since then and resulted in the BRPerfX software. This tool was applied to the REFAP's HGU last year and the presentation of this project and its results is the main concern of this paper.

In the sequence of this work we describe the HGU and the BRPerfX software in section 2. In section 3, the main results are presented. These results encompass the project main steps, the development of the base case, the actions made to improve the dynamic performance, the evidences of that improvement, and the estimation of the economic gains after the actions made by our technical team. Finally, we present some brief concerns about the overall work in the conclusions section.

2. Materials and Methods

2.1. HGU Description

The REFAP HGU started up in December 2005 and produces hydrogen up to 99.9 % purity level from natural gas. The unit objective is to supply hydrogen for the subsequent diesel hydro treating unit according to its demand. As the diesel is the main REFAP's product and it can only be sold after being hydro treated, the continuous and stable operation of HGU is fundamental for attending one of the most important strategic enterprise objectives. The HGU is composed by three main stages (Fig. 1):

2.1.1. Desulfurization and Pre-heating

In this stage, the natural gas is fed in the unit and receives a hydrogen injection, being immediately sent to heating in the convection section of the reformer furnace. There, it acquires the necessary temperature to start up the reactions.

2.1.2. Conversion

Here, the pre-heated load receives steam and is converted to CO, CO_2 and H_2. After, the final conversion from CO to CO_2 and H_2 is carried out in the shift reactor.

2.1.3. Purification (PSA – Pressure Swing Adsorption)

The PSA purpose is to purify the hydrogen produced in the HGU until it reaches a concentration of 99.9 % (volumetric). It utilizes the molecular sieve process. The gas is admitted at the bottom of the vessel where the impurities filtering and retention occurs.
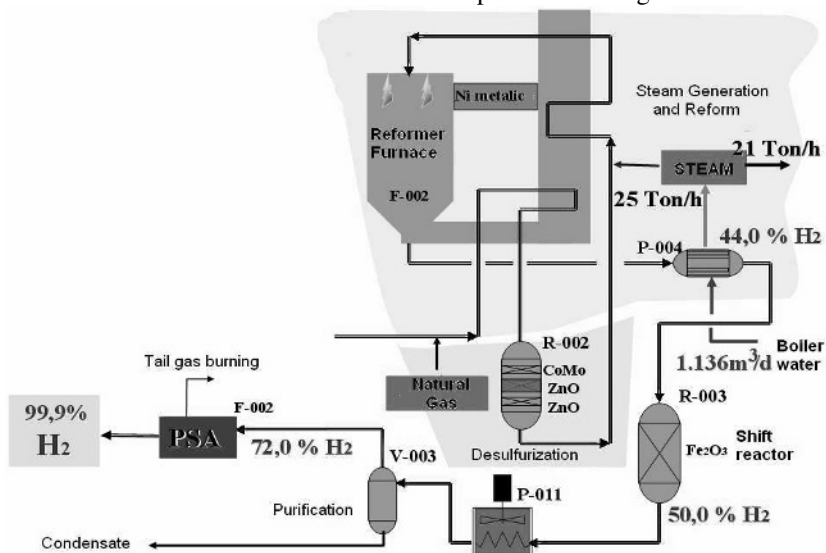
Fig. 1: HGU scheme. The HGU produces 550,000 Nm3/d of H2.

2.2. The BRPerfX tool

The BRPerfX is a real-time tool for performance monitoring and evaluation of control loops developed by Federal University of Rio Grande do Sul and Trisolutions enterprise and supported by PETROBRAS. It allows the identification of the main responsible for the bad control performance of a plant with large number of control loops. This is possible due to the state of art level of indexes and metrics available in the software. These includes: (1) oscillation detection indexes, (2) performance metrics based on minimum variance index, (3) traditional statistics, (4) Valve analysis, (5) controllers service factors, (6) PID tuning changing monitoring, (7) several graphical analysis and, (8) customized weighted grades for equipments, units and plants.

The application works online without any interference in the plant operation. The essential information required for its correct work are the PV (Process Variable), SP (Set-Point), CO (Controller Output) and, MO (Operation Mode). To get these data, it uses an OPC (OLE for Process Control) DA (Data Access) interface, a standard to connectivity among systems largely adopted in industry. The software architecture is composed by three main parts: a Windows service, a database and, a Web application. A typical view of the web user interface is shown if Fig. 2
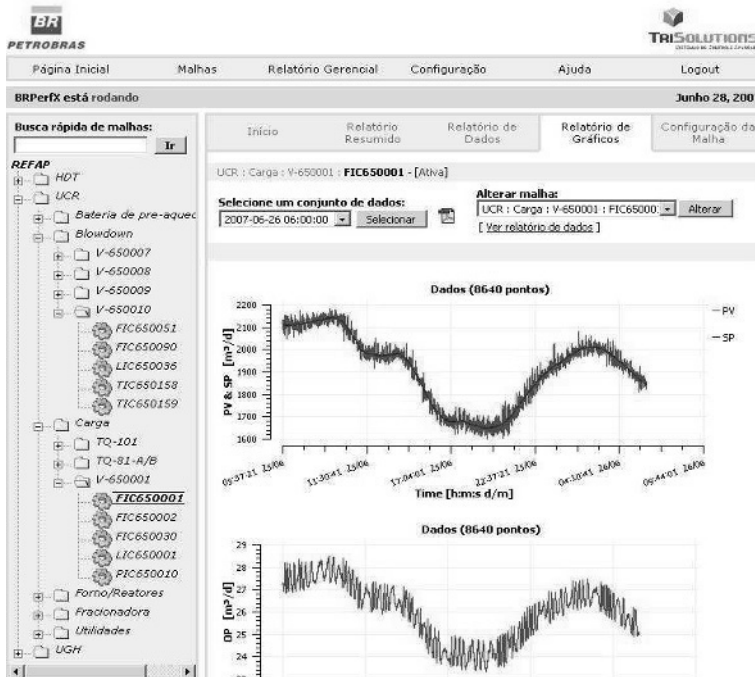
Fig. 2: BRPerfX user interface in web environment.

3. Results

3.1. Project Stages

To be successful, an industrial control management project should be composed by four basic steps plus cycles for continuous improvement. The time ranges, estimated in the following lines, are the ones for a typical unit with about 50 to 100 control loops. These steps are: (1) software installation and configuration – this usually takes one week. For the HGU case it was very fast, since the software was already installed. (2) control loop

evaluation and elaboration of a base case – at least one month, depending on plant operating conditions. This stage is important because it furnishes the reference for future performance comparisons, (3) Corrective actions – the range time depends on problem's nature. The time to solve tuning problems can be estimated about 10 loops for week. Maintenance and project problems have much slower solving velocity. (4) Process revaluation and economic gains estimate – this will show the relevance of the work for the high level enterprise staff. It can be estimated about, at least, one month to collect data and one week to compute the gains. This will be explained in section 3.5.

All these steps must be followed by new cycles for continuous improvement and preventing performance degradation.

3.2. The Base Case

The base case for this work was done with data collected from 2008 August 8 up to 2008 September 8. In this time range the unit was under stable operation and no actions were made by the control engineers. After a carefully analysis of data, the base case was defined as the one briefly presented in Table 1.

Table 1 – The base case

Number of loops:	42	
Problem classification	#loops	Percent
DCS configuration	2	5%
Other	3	7%
Maintenance	1	2%
Tuning	21	50%
Operation	4	10%
Project	1	2%
No problems detected	10	24%
Saturation (>= 25% of time)		
Loops	13	31%
Operation mode (>=25% of time in manual)		
Loops	9	21%

3.3. Actions Made

The actions made by the control engineers were focused mainly in tuning and DCS configuration problems. A new cascade controller was also included after a review in the HGU control philosophy. This work took about two discontinuous months. The results, in comparison with the base case, are presented in Fig. 3.
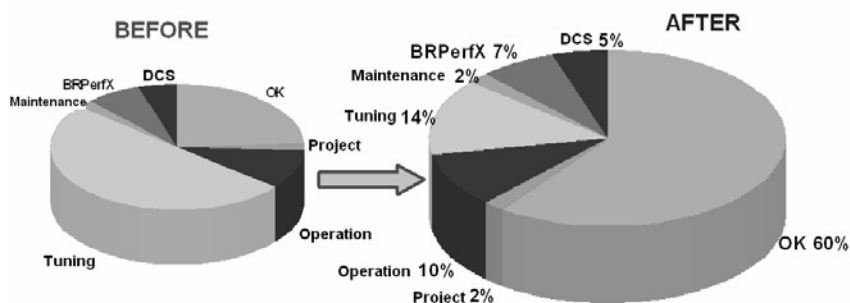
Fig. 3: Evaluation of control loops before and after the actions made by the control engineer team.

3.4. Dynamic Improvement

The process improvement can be clearly seen by looking some pictures available in the BRPerfX software. Due to space limitations a small fraction of them are presented in this section. Some parts of the graphics are written in Portuguese (software language).

The first example concerns the O_2 excess in the reformer furnace, controlled by the controller AIC702002. This controller is the master of a cascade that operates in the air fan blower. This controller had a very conservative performance, being very slow and oscillating. As a consequence, the operators had orders to operate at a higher than recommended oxygen excess. After the control engineers work, the standard deviation of the loop goes down from 0.41 to 0.06 (Fig. 4). This also created a potential to operate at lower levels of air excess. Despite being a small furnace, this allows a fuel gas reduction and a monthly economic gain about R$ 10.000,00, or USD 4,400.00 / month (using the 2009 February, 6 exchange rates).
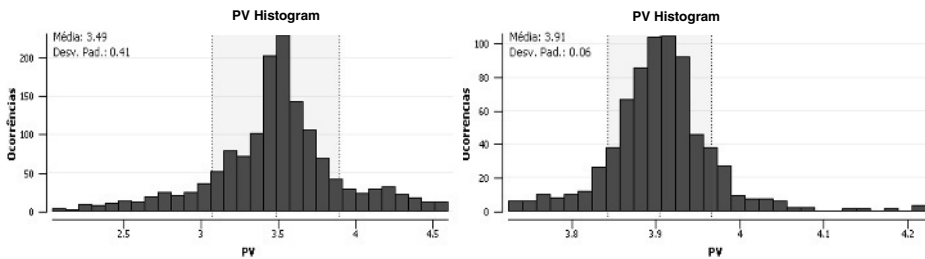
Fig. 4: BRPerfX histogram of AIC702002 before and after the control engineers actions.

The second example is the pressure control of the reformer furnace pilot burner, PIC702051. After our action, its very high oscillations have vanished, as Fig 5 shows.
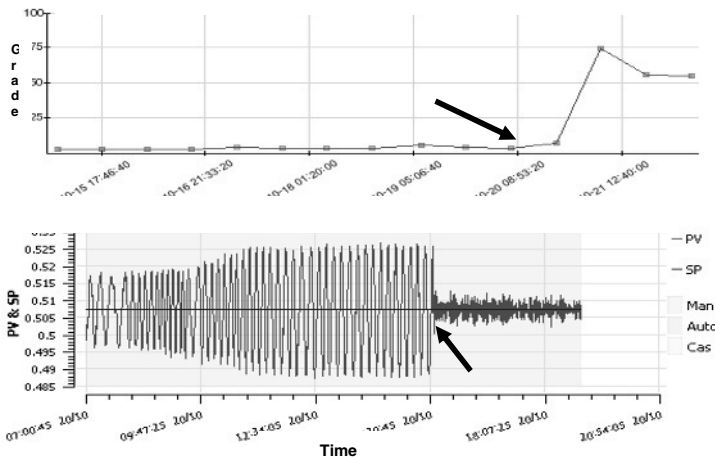
Fig. 5: Evaluation of PIC702051. In the top, the one week control loop grade evolution. Below, it is shown the PV-SP graphics, for a 12 hour period. The arrows point the time we have acted in.

Another illustrative example of the insights obtained by using the software occurred in the natural gas recycle pressure controller, at the feed HGU section – PIC702003. This controller presented very high oscillations with nearly 2 minutes period, as can be seen in the power (Fourier) spectra computed by the software (Fig. 6). After retuning, the

performance improvement - also shown in Fig. 6 – is obvious. The spectra became more distributed and less dense in the higher frequencies.
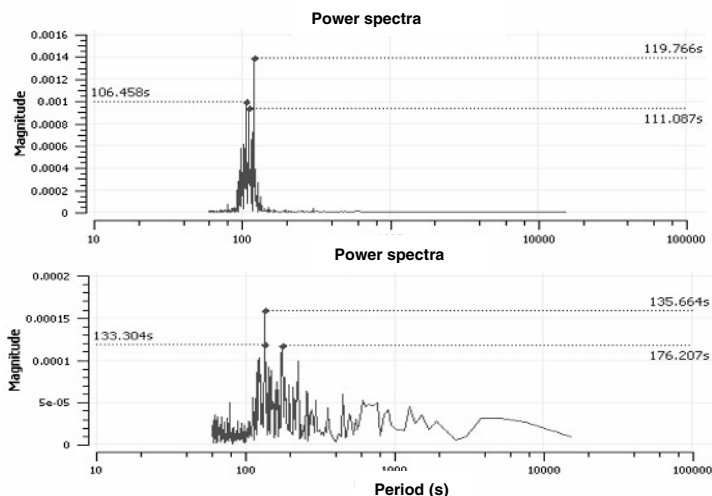
Fig. 6: Power spectra after (top) and before (bottom) the control loop performance improvement.

More interesting than showing the better performance, the new index (TuniPID) [2] included in the software indicates to the user that the control can be improved, but not by a PID because the control loop is very close to the best achievable PID performance.

3.5. Economic gain

Despite showing the dynamic behavior is much better now than ever in the HGU short history; this has no significant meaning for the enterprise managers and maybe for other people but control engineers. So, it is fundamental to compute the economic gains of the whole work. This computation varies from one unit to another and there is no applicable methodology avoiding a process analysis. Facing this fact, REFAP's and Trisolutions engineers internally developed a framework to compute gains from regulatory control improvement for chemical plants. This is based on comparisons with base case performance, accounting the price evaluation (for reactants, intermediate products and fuels) and, the formalization of the basic tangible mechanisms for plant profit: selectivity, quality, productivity, fuel efficiency, loss reduction, among others. After one week of careful calculations, we estimate the official number for the economic gain of this work about USD 85,000.00 per month, plus the intangible gains related to plant stability and the consequent reduction of scheduled and/or unplanned plant shutdown.

4. Conclusions

This work presents a project that applies new academic developments to solve practical industrial problems. Due to space limitations, we conclude only stating that the critical factors to its successful implementation were: IT infrastructure, reliable tool choice and, staff and management commitment. We thank to all refinery staff for their gentle help.

References

[1] Harris, T. J., 1989, Assessment of control loop performance, The Canadian Journal of Chemical Engineering, 67, p. 856-861.
[2] Farenzena, M., 2008, Novel Methodologies for assessment and diagnostics in control loop management, Ph.D. Thesis, Federal University of Rio Grande do Sul.

Advanced Control and Optimization of a Natural Gas Plant - Benefits of the New Regulatory Control Strategy

Mario C.M. Campos,[a] Marcos V. C. Gomes,[b] Vicente D. Moreira,[c] Marcos F. Lima,[d] José Renato Silva,[d]

[a]PETROBRAS / CENPES / Engenharia Básica/ Automação, Equipamentos Dinâmicos e Confiabilidade, Av. Horácio Macedo, 950 - Ilha do Fundão, Rio de Janeiro, 21949-915, Brazil (e-mail: mariocampos@petrobras.com.br).
[b]PETROBRAS / CENPES / P&D de Gás, Energia e Desenvolvimento Sustentável / Gás Natural / Célula de Otimização
[c]PETROBRAS / UN-RNCE / ENGP, Natal, Brazil
[d]PETROBRAS / UN-RNCE / UTPF, Natal, Brazil

Abstract

This paper presents the design of a new regulatory control strategy, as part of an advanced control and optmization project, for the natural gas processing plant of PETROBRAS (UPGN-III in Portuguese symbols) in Rio Grande do Norte. This phase of the implementation of the advanced control and optimization system (multivariable model based constraint control) allowed an increase of about 4% in the Liquefied Petroleum Gas (LPG) per charge and composition, this is the principal product of this unit, representing an income of US$ 1.300.000 a year. The new control system also reduced in 30% the number of unscheduled unit's downtime due to an improvement of its stability.

Keywords: advanced control system, regulatory control, industrial application, loop performance assessment.

1. Introduction

The advanced control system in petrochemical plants is an industrial reality (Qin and Badgwell,, 2003). There are many advantages for the process units in using these systems as more stability, respect to the restrictions and an increase in the profitability and safety. PETROBRAS has been investing in the development of these systems for several years. It is a consolidated technology in its refineries with many controllers for FCC and distillation units. However, the application in Natural Gas Plants (UPGNs in Portuguese symbols) is recent, although the optimization of these plants can bring great economical earnings, besides to increase the energy efficiency and to minimize the emissions.

Installation and maintenance of advanced controllers with a good performance is a big challenge. Their performance is influenced by instrumentation problems, bad tuning of the regulatory control, bad identification of the dynamic models of the process (Ender, 1993)(Kern, 2007), no measured disturbances, etc. In this article, we will discuss some tools that were used to the diagnosis and tuning of the regulatory control and the instrumentation, which is an important phase of the implementation of an advanced control in a Natural Gas Plant. We also show some results obtained in a real case during this phase.

2. Advanced Control System

Process control aims to maintain certain variables in their desirable operational limits and could be visualized as a pyramid. In the base of this pyramid, the first level is the regulatory control, that uses PID controllers (Campos and Teixeira, 2006)(Ogata, 1982) and is configured in the digital systems (DCS - Distributed control system or PLC - Programmable logical controllers). In a second level, we have the advanced control systems that use for instance Model Predictive Control (MPC). This algorithm considers the interaction between control loops, and includes an optimization layer of the industrial plant. These algorithms are usually implemented in a process computer that communicates with the DCS using for example OPC protocol (OPC, 2008). The outputs of this advanced control are usually the setpoints of the PID controllers of the regulatory control. So if there is a problem in the advanced control, the plant operation continues with the last PID setpoints in the DCS.

Model predictive control tries to control several process variables simultaneously acting in many manipulated variables. The advanced control algorithm used in this natural gas plant is called CPM (SSP-Laplace) and was developed by PETROBRAS/CENPES.

CPM algorithm incorporates a dynamic model of the process, used to obtain the control actions that minimize the effects caused by plant disturbances. The multivariable and dynamic model of the process allows to increase the performance of the control system mainly when we have some of the following problems:

- Multivariable and coupled control loops;
- Process with long dead times or long time constants.

Best operating setpoint is obtained by an optimization module of the CPM. This module determines this point based on economical criteria, considering the process constraints.

An advanced control system won't reach the expected economical benefit if turned off constantly by the operators. Therefore, the instruments, valves and the regulatory control (PIDs) should operate appropriately. Besides, training is also a key factor to the technology absorption for the plant personnel (operators and process engineers), in order to use the system adequately. The phases to implement an advanced control system are: Functional project (select the economical objectives, the controlled and manipulated variables and disturbances), Diagnosis, evaluation and tuning of the regulatory control and instrumentation; Identification of the process models; Installation and tuning of the model predictive control and Training, maintenance and documentation;

In this article we will focus in the results of the phase associated with the diagnosis, evaluation and tuning of the regulatory control and instrumentation. The project of advanced control for the natural gas plant (UPGN-III) of Guamaré - Rio Grande do Norte (Northeast of Brazil) began in March of 2007.

3. Description of the Natural Gas Plant (UPGN-III)

The natural gas plant of PETROBRAS/UN-RNCE (UPGN-III) recoves the heavy fractions (LNG) contained in the natural gas. First, this gas is chilled and after the expander its temperature is enough to condense the heavy hydrocarbon fractions (C_{3+}). The flow is sent to a distillation column that separates the gas in the top, known as residual gas, from the liquid (LNG). The residual gas exchanges heat with the feed in a cold box. Afterwards, it goes to the suction of the turbo-expander compressor that is in series with the sale gas compressor. Economically the objective is obtain minimum lost of C_{3+} in this gas (Campos et al., 2008).

4. Diagnosis, evaluation and tuning of the regulatory control

The good performance of the regulatory control is fundamental to the success of the advanced control system. This is due to the fact that the advanced system acts usually changing the regulatory setpoints. Therefore, if a problem exists in the regulatory control then the performance of the advanced system is affected.

Common problems associated with the regulatory control are: tuning of the PID controllers (oscillation, stability, etc.), control strategy (interaction, degrees of freedom, etc.), and instrumentation problems (out of scale, measurement resolution, sizing, etc.).

So, it is clear the importance of the diagnosis and evaluation of the regulatory control. To help us in this task, reducing the subjectivity, PETROBRAS developed with Federal University of Rio Grande do Sul (UFRGS in Portuguese symbols) the software "BR-PerfX". It calculates some key performance indicators and generates graphs that are used to quantify the quality of the regulatory control (Kempf, 2003)(Farenzena and Trierweiler, 2008). From the analysis of these results it is possible to identify the problem affecting the regulatory control.

If the problem is the tuning of PID controller then the software "BR-Tuning" is used. This software was developed from a partnership between PETROBRAS and the Federal University of Campina Grande (UFCG in Portuguese symbols).

BR-Tuning (Schmidt et al., 2008)(Arruda and Barros, 2003) implements a group of techniques for open and close loops identifications and proposes new tuning parameters. It communicates directly with the process automation system (DCS or PLC) using the OPC protocol (OPC, 2008).

The main problem found in this plant (UPGN-III) was the turbo-expander flow control (FIC-01) operating continuously in manual mode. When the control was putted in automatic mode occurred a strong interaction with the pressure control (PIC-01), resulting in oscillations. It was the cause of many trips of the unity, and was a great operators' concern. Figure 1 displays the old control strategy of the unit.
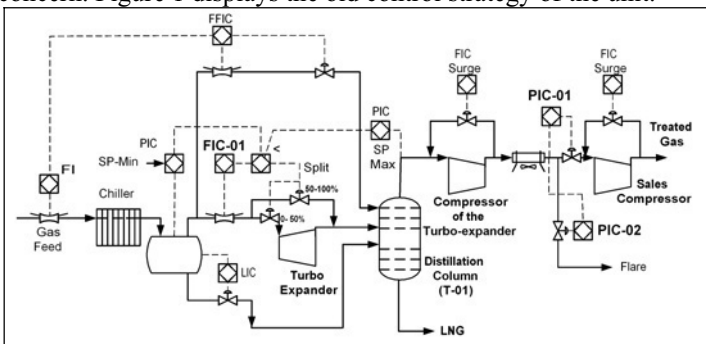
Figure 1 – Old regulatory control strategy.

The problems with the old control strategy were the following ones:

- The setpoints of the pressure controls in the suction of the compressor associated with the turbo-expander (PIC-01 and PIC-02) were difficult to be adjusted, because they should be changed if the turbo-expander was on or off. For example, if there were a trip in the turbo-expander, these setpoints should be changed quickly to lower values in order to avoid a trip in the whole unit due to high pressure in the distillation column (T-01). This trip was observed many times in this plant.

- The degree of freedom was not well used, because the control strategy was fixing many variables of the compressor associated with the turbo-expander. Therefore the control system was not stable in automatic mode.
- When the compressor associated with the turbo-expander opened its anti-surge recycle valve (Campos and Teixeira, 2006), the discharge pressure fell and the controller (PIC-01) closed the suction valve of the sales compressor, taking in a not correct manner this machine also to surge. This event took some times the unit to oscillate and to trip, generating losses.
- The control pressure of the distillation column (T-01) didn't operate in automatic mode in spite of being the most important column of the plant. Abrupt variations in this pressure increase the losses of LNG.

A new control strategy was proposed, according to figure 2. Basically, we changed the location of the process variables for the pressure control (PIC-01 and PIC-02). These controllers are now monitoring the pressure of the distillation column (T-01).
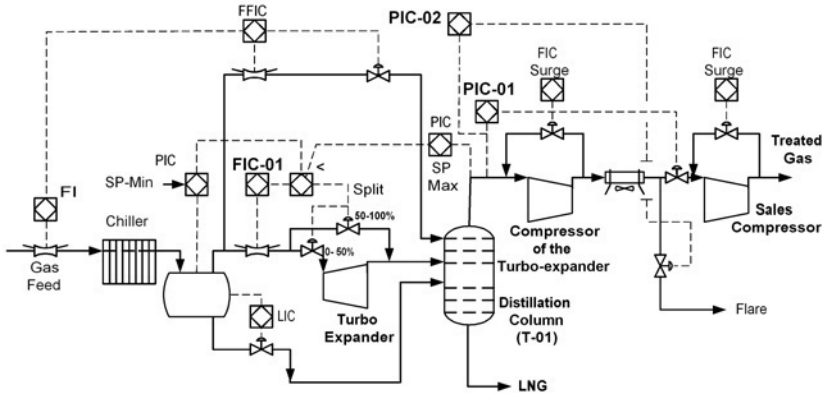
Figure 2 – New regulatory control strategy to UPGN-III.

The turbo-expander flow control (FIC-01) in this new strategy will define the compressor speed but the discharge pressure is free to accommodate changes in compressors efficiency. This new regulatory control strategy does not the old problems associated with the unit trip. Another advantage is that pressure control is now in automatic with a good performance (figure 3), manipulating the suction valve of the sales compressor (PIC-01).

The plant flow control (FIC-01), that manipulates the gas to the turbo-expander, is now operating in automatic mode continually (figure 4: FIC-01 performance - setpoint: green curve, flow: blue curve, and the valve opening: yellow curve). The next phase of the advanced control project will be the model identification and the tuning of the model predictive controller. This new phase will be discussed in a next paper.

5. Economical benefits and safety result in this project phase

With this new regulatory control strategy, the unit is easier to be operated and it is easier to find a better operating point, allowing to increase the profitability of the plant. Besides, more than fifty control loops were analyzed and tuned. Many levels controls were tuned in order to minimize the movement of the associated flow, improving the stability.
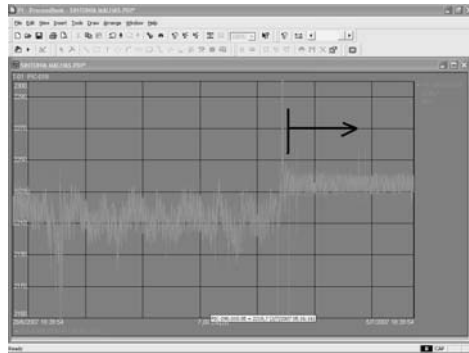
Figure 3 – Greater stability of the distillation column pressure.
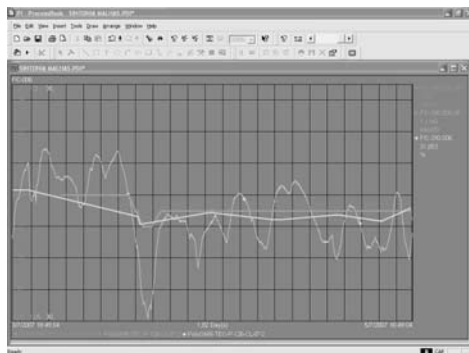
Figure 4 – Flow control in automatic, manipulating the gas to expander.

LNG of the T-01 (figure 1) is sent for another distillation column (not shown in figure 1) where Liquefied Petroleum Gas (LPG) is obtained, which is the principal product of this unit (UPGN-III). In order to calculate the benefits of this project, we defined a performance indicator based on the LPG production. This indicator considers the production of LPG in relation to the C_{3+} in the feed gas:

$$Indicator = \frac{LPG\ Flow}{Feed\ Flow} * \frac{1}{\%\, C_3^+ \text{ in the Feed}}$$

After the tuning of this new control strategy, this indicator passed from 3,17±0,17 to 3,30±0,08. So it was obtained an increase of about 4,1% in the production of LPG. This result represents an increase in the profit of the unit of approximately US$ 1.300.000/year. We can also observe a reduction in the variability of the plant (smaller standard deviation). Figure 5 shows the performance indicator improvement after the implementation and tuning of the new control strategy. Another benefit is associated with the reduction of the number of trips (non-programmed stops) of the unit. The medium time of the unit stopped passed from 729 minutes a month (May of 2006 to March of 2007) to 485 minutes a month (July of 2007 to May of 2008). This means a reduction of about 33%. It also should be noticed that less trips implicates in less emissions. So, the system also improved the health, safety and environment aspects (HSE) of the plant.

6. Conclusions

This paper presented the results obtained with the implementation of a new regulatory control strategy, which is an important phase to the implantation of the advanced control

system. It showed that the old control strategy was unstable and responsible for many unit trips. A new regulatory control strategy was able to solve these old problems. The next phases of this project will be the identification of the models and the tuning of the predictive controller.
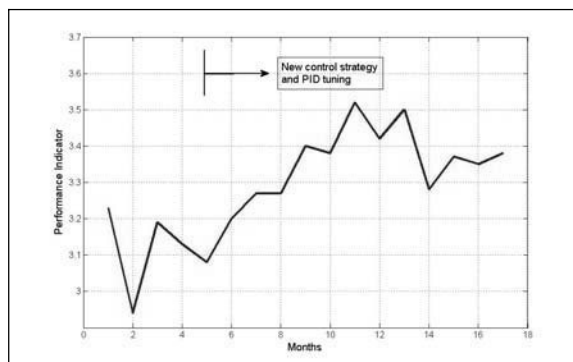
Figure 5 – Better performance indicator after the implementation of the new control strategy.

However, it is worthwhile to emphasize the economical results obtained with the implantation of this new regulatory control strategy. Only this phase allowed an increase of approximately 4% in the LPG production (an increase in the profitability of US$ 1.300.000/year). Other observed benefit was an expressive reduction of 33% in shutdown time of the unit.

Acknowledgements

References

Arruda, G. and Barros, P., 2003, "Relay based gain and phase margins PI controller design", IEEE Transactions on Inst. And Meas. Tech., 52(5), 1548-1553.

Campos, M. and Teixeira, H., 2006, "Controles típicos de equipamentos e processos industriais", Ed. Edgard Blücher, São Paulo.

Campos et al., 2008, "Sistema automático de identificação de modelos dinâmicos necessários ao Controle Avançado – Aplicação na UN-RNCE", Primeiro CICAP – Congresso de Instrumentação, Controle e Automação da PETROBRAS, May, Rio de Janeiro.

Ender, D., 1993, "Process Control Performance: Not as good as you think", Control Eng., 1993, pp.180.

Farenzena, M. and Trierweiler, J., 2008, "Fronteiras e desafios em gerenciamento de malhas de controle", In: COBEQ 2008 - Congresso Brasileiro de Engenharia Química, Recife.

Kempf, A., 2003, "Avaliação de Desempenho de Malhas de Controle", Dissertação de Mestrado, Departamento Eng. Química, Universidade Federal do Rio Grande do Sul, UFRGS.

Kern, G., 2007, "Summiting with multivariable predictive control", Hydrocarbon Processing,

Ogata, K., 1982, "Engenharia de Controle Moderno", Ed. Prentice/Hall do Brasil.

OPC Foundation, 2008. Site: <http://www.opcfoundation.org/>, Accessed in: 03/17/08.

Qin, S. and Badgwell, T., 2003, "A survey of industrial model predictive control technology", Control Engineering Practice 11, 733-764.

Schmidt et al., 2008, "BR-Tuning Ferramenta para sintonia de controladores PID", Primeiro CICAP – Congresso de Instrumentação, Controle e Automação da PETROBRAS, May, Rio de Janeiro.

1689

Versatile Biodiesel Production by Catalytic Separative Reactors

Anton A. Kiss

AkzoNobel Research, Development & Innovation, Process & Product Technology, Arnhem, The Netherlands, tony.kiss@akzonobel.com

Abstract

This study proposes an integrated biodiesel production via a two-step process that combines the advantages of using solid acid and base catalysts with the integration of reaction and separation. Such an integrated separative reactor is flexible to treat any range of free fatty acids present in the fatty raw material. Computer aided process engineering tools such as AspenONE are used for process design and simulation of a plant producing 10 ktpy biodiesel from animal fat and bio-ethanol.

Keywords: reactive distillation, green catalysts, solid acid / base, biofuels

1. Introduction

The recent steep increase in fossil fuel prices associated with governmental restrictions on discharge of green-house gasses shifted the worldwide trend to focus on renewable energy sources. Biodiesel is a very popular renewable fuel, currently produced from vegetable oils, animal fat or even recycled waste cooking-oil from the food industry.[1,2] Due to its properties similar to petrodiesel, biodiesel can be used in pure form, or may be blended with petroleum diesel at any concentration, in most modern diesel engines. As a *green fuel*, biodiesel that has many advantages over conventional petrodiesel: it is safe, renewable, non-toxic and biodegradable, it contains insignificant amounts of sulfur and its increased lubricity extends the life of diesel engines. In addition, it has a high cetane number (above 60 compared to 40 for petrodiesel), a high flash point (>130°C) and it emits ~70% fewer hydrocarbons, ~80% less CO_2, and ~50% less particles.[1-3]

Biodiesel is a mixture of fatty esters, currently produced by (trans-)esterification of tri-glycerides and free fatty acids, followed by several neutralization and purification steps. However, all the traditional methods suffer from drawbacks related to the use of liquid acid/base catalysts, heading to major economical and environmental penalties, especially considering the recent boost of the international biodiesel production rate. Worldwide, the production of biodiesel increased tremendously during the past 10 years, mostly in Asia, US, and Western Europe with Germany, France, Austria, Spain and UK among top consumers (Figure 1).

This work proposes a novel two-step biodiesel production process bases on reactive-separation using solid acid/base catalysts, thus simplifying the overall process and bringing significant benefits: high conversion and selectivity, elimination of conventional catalyst-related operations, no waste streams, as well as reduced capital investment and operating costs. The process design of the 10 ktpy fatty acid ethyl esters plant described here is based on experimental results, integrated in rigorous simulations performed using AspenTech AspenONE as computer aided process engineering tool.
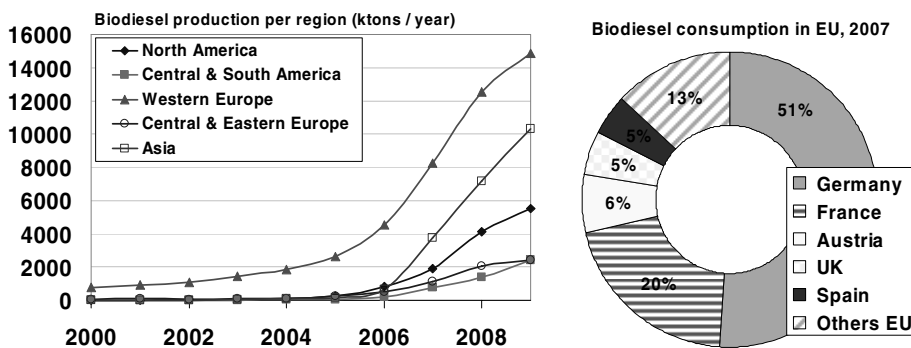
Figure 1. Biodiesel production per region (left), and biodiesel consumption in EU (right).

2. Problem statement

All conventional biodiesel production methods have associated optimal operating parameters and downstream processing steps, although much of the available literature emphasizes the base catalyzed route.[1-3] Traditional biodiesel processes employ liquid catalysts, such as H_2SO_4 or KOH.[3] The problem is that such catalysts demand neutralization, separation, washing, recovery, and salt waste disposal operations with serious economical and environmental consequences.

Nowadays, the surplus of waste oil available at industrial scale would allow production of very cheap biodiesel – a key benefit in the energy market. For example, in Brazil alone, more than 350 millions litres of biofuel are produced annually from animal fat. The problem with the animal fat or waste-oil, is that it becomes useless within 24 hours since it turns so acidic due to the increased free fatty acids (FFA) content, that it is more appropriate for making soap than for biodiesel. To solve these problems, we propose a sustainable two-step process based on the esterification of FFA's in a separative reactor using solid acids,[4,5] followed by trans-esterification of the remaining tri-glycerides (TG) using conventional or solid base catalysts.

$$\textbf{1}: \text{R-COOH} + \text{EtOH} \leftrightarrow \text{R-COO-Et} + H_2O \quad \text{(esterification)}$$

$$\textbf{2}: \text{TG} + 3\,\text{EtOH} \leftrightarrow 3\,\text{RCOO-Et} + \text{Gly} \qquad \text{(trans-esterification)}$$
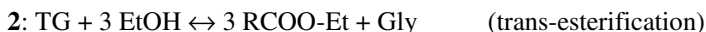
The integrated reactive distillation equipment proposed in this work is able to shift the chemical equilibrium and drive the esterification reaction to completion by continuously removing the fatty esters products and water by-product.[6,7] The raw materials consist of waste-oil or animal fat – mainly a mixture of free fatty acids – and a light alcohol, such as methanol or (bio-)ethanol. A key feature of this work is the replacement of anhydrous ethanol by its hydrous azeotrope, thus leading to further reduction of production costs.

Table 1 presents an overview of the available solid acid and base catalysts for biodiesel production by (trans-)esterification.[4-6] In this work we selected the metal oxides as solid acid catalysts for FFA esterification (first step) and calcium ethoxide as solid base catalyst for the trans-esterification of the remaining tri-glycerides (second step).

Table 1. Advantages and disadvantages of the acid/base catalysts tested for (trans-)esterification.

Catalyst type	Benefits	Drawbacks
Ion-exchange resins (Nafion, Amberlyst)	Very high activity Easy regeneration	Low thermal stability Possible leeching
TPA ($H_3PW_{12}O_{40}$)	Very high activity	Soluble in water
TPA-Cs ($Cs_{2.5}H_{0.5}PW_{12}O_{40}$)	Super acid sites	Low activity per weight
Zeolites (H-ZSM-5, Y and Beta)	Controlable acidity and hydrophobicity	Small pore size Low activity
Sulfated metal oxides (zirconia, titania, tin oxide)	High activity Thermally stable	Deactivates in water, but not in organic phase
Niobic oxide (Nb_2O_5)	Water tolerant	Average activity
Calcium oxide / CaO	Low temperatures	Long reaction times
Calcium methoxide / $Ca(OMe)_2$	High yield, reusable	High reactants ratio
Calcium ethoxide / $Ca(OEt)_2$	High yield, short times	High reactants ratio
Li-dopped zinc oxide / ZnO	Low temperatures	Long reaction times
KF loaded on Eu_2O_3	Short reaction times	Incomplete yields

3. Simulation methods

The simulation methods available are given in Table 2. Each method has important benefits but also certain drawbacks and the requirements can differ significantly. The amount of data required by the rigorous method is practically not feasible in practice while the shortcut method leads to low-fidelity models only, with limited applications.
For practical reasons, the hybrid approach gives the best results. In this work the experimentally determined kinetic parameters were used [5,6] but the fatty components were lumped into one fatty acid/ester compound, according to the reaction:

$$R\text{-}COOH + EtOH \leftrightarrow R\text{-}COO\text{-}Et + H_2O$$
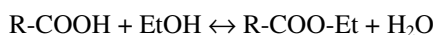
Table 2. Simulation methods for biodiesel production: requirements, benefits and drawbacks.

	Rigorous method	Shortcut method	Hybrid method
Requirements	Properties for all species. VLL data and BIP's for all pairs of components. Kinetic parameters for all reactions possible.	Properties for single fatty acid/ester/tri-glyceride. VLL data for the system ester/glycerol/alcohol. Asumed conversion (no kinetic parameters).	Single or reduced list of fatty acid/ester/TG. Short list of VLL data and BIP's for components. Reduced list of kinetic parameters, few reactions.
Benefits	Easy optimization of reaction and separation. High fidelity model. Usable for many plants. Easy comparison for various feedstocks.	Simple model. Fast simulations. Easy-to-build mass and energy balance. No data needed for all species present.	Optimization possible for reaction and separation. Certain ability to compare various feedstocks. Better model fidelity. Fast simulations for RTO.
Drawbacks	Slow simulations and convergence problems. Expensive measurements. Limited RTO and model based control usage.	No comparison possible for various feedstocks. Low-fidelity model. Less ability to use RTO.	More effort to build component list and get kinetic parameters. More work to find VLL data and regress BIP's.

4. Results and discussion

The properties of the fatty components were determined experimentally, or estimated using state-of-the-art contribution methods such as UNIFAC – Dortmund modified.[5-9]

Figure 2 (left) shows the residue curve map (RCM) for the ethanol-water-glycerol ternary mixture. The presence of the ethanol-water azeotrope does not hinder the biodiesel production process, since ethanol is a reactant and not a high purity product. Moreover, as water by-product from the esterification reaction further dilutes the ethanol, its hydrous azeotrope can be used directly as a lower-cost feedstock.

Vapor pressure is perhaps one of the most important properties with a critical effect in modeling reactive separations. Figure 2 (right) shows the vapor pressure of most common fatty acids and esters. At ambient pressure the boiling points are relatively high, exceeding 300 °C. Although high purity products are possible by reactive distillation, the high temperature in the reboiler – caused by the high boiling points, is in conflict with the thermo-stability of the biodiesel product. However, this problem can be avoided by working at lower pressure or by allowing ethanol in the bottom product.

Figure 3 presents the flowsheet of a two-step biodiesel production process based on a reactive distillation column (RDC) as the key unit for esterification or pre-treatment of free fatty acids (FFA). The process proposed here was rigorously simulated and optimized using AspenTech AspenONE. The production rate considered for the plant designed in this work is 10 ktpy fatty acid ethyl esters (FAEE). Note that the kinetic parameters used in the simulation were previosly reported in the open literature.[5,6]

The RDC is operated in the temperature range of 100–250 °C, at ambient pressure. Out of the 15 stages of the integrated unit, the reactive zone is located in the middle of the column (10 stages). The fatty acid is pre-heated then fed as hot liquid on top of the reactive zone while a stoichiometric amount of alcohol introduced in the bottom of the reactive zone, thus creating a counter-current V-L flow regime over the middle reactive section. The reflux ratio is very low (RR=0.1) as returning water to the column is detrimental to the chemical equilibrium. Water by-product is removed in top, then separated in a decanter from which only the fatty acids are recycled to the column while water is recovered at high purity and hence reusable as industrial water on the same site.
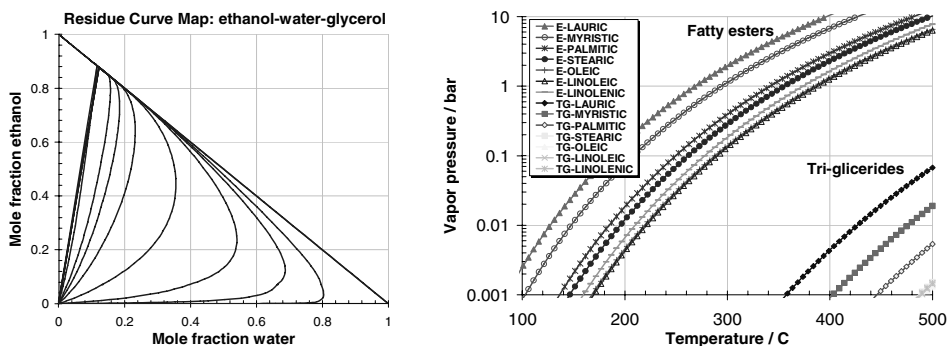
Figure 2. RCM ethanol-water-glycerol (left), Vapor pressure of fatty esters vs temperature (right).
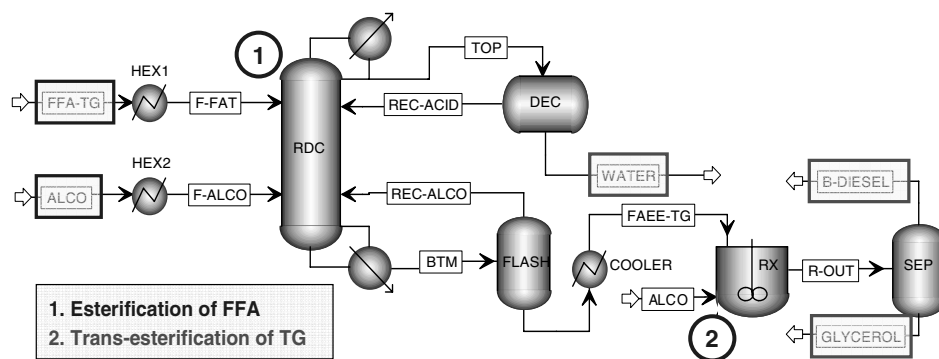
Figure 3. Flowsheet of biodiesel production by catalytic reactive distillation.

The fatty esters are delivered as high-purity bottom product of the RDC. The hot product is flashed first to remove the traces of ethanol, then it is send to the trans-esterification reactor to further convert the remaining tri-glycerides to fatty esters. In this work we considered the worst case scenario, with 100% FFA in the feedstock.

The mass and energy balance is given in Table 3. High purity products are possible, the purity specifications exceeding 99.9%wt for the final biodiesel product (FAEE stream). Note that the total amount of the recycle streams (REC-ACID and REC-ALCO) is not significant, representing only ~0.5% of the total biodiesel production rate.

Figure 4 (left) shows the liquid composition profiles in the reactive distillation column. The concentration of fatty acids decreases while the concentration of fatty esters increases from the top to bottom. Similarly, the ethanol concentration decreases while water concentration increases from bottom to top. The temperature and reaction rate profiles in the RDC are presented in Figure 4 (right). As expected, the reaction rate exhibits a maximum in the middle of the column, in the reactive zone. Moreover, the concentration of water is low in the reactive zone, hence the catalyst activity is not affected. Nevertheless, the concentration of reactants is relatively high and the temperature is sufficiently high to allow high reaction rates and complete conversion.

Table 3. Mass and energy balance of a 10 ktpa biodiesel production process, based on RD.

	F-ACID	F-ALCO	BTM	TOP	REC-ACID	REC-ALCO	FAEE	WATER
Temperature K	418.1	352.2	507.9	372.7	303.1	372.6	303.1	303.1
Pressure atm	1.036	1.036	1.017	0.987	1	1	1	1
Vapor Frac	0	0	0	0	0	0	0	0
Mass Flow kg/hr	1094.918	264.202	1257.312	109.246	0.152	7.312	1250	109.094
Volume Flow l/min	22.986	6.156	28.953	1.983	0.003	0.154	23.774	1.839
Enthalpy Gcal/hr	-0.892	-0.394	-0.886	-0.405	0	-0.006	-1.035	-0.413
Mass Flow kg/hr								
ETHANOL	0	253.568	2.622	0	0	0.857	1.765	0
ACID	1094.918	0	0	0.115	0.107	0	0	0.008
WATER	0	10.635	0.029	109.089	0.003	0.013	0.016	109.086
ESTER-E	0	0	1254.661	0.043	0.042	6.442	1248.219	0
Mass Frac								
ETHANOL	0	0.96	0.002	0	0	0.117	0.001	0
ACID	1	0	0	0.001	0.702	0	0	0
WATER	0	0.04	0	0.999	0.02	0.002	0	1
ESTER-E	0	0	0.998	0	0.278	0.881	0.999	0
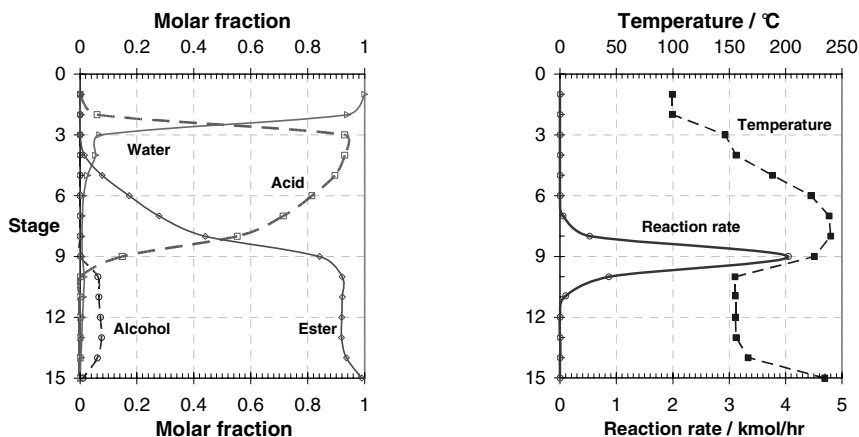
Figure 4. Profiles in RDC: liquid composition (left), temperature and reaction rate (right).

5. Conclusions

An innovative two-step biodiesel process based on separative reactors using solid catalysts was developed in this study using computer aided engineering tools such as AspenTech AspenONE. The novel two-step process proposed here improves considerably the biodiesel production and reduces drastically the number of downstream processing steps. The major benefits of this sustainable process are:

- Flexible integrated reactor suitable for a large range of fatty raw material with up to 100% FFA content, such as: frying oils, animal tallow, tall oil, waste vegetable oil.
- Straightforward and robust process with no soap formation, no catalyst-related waste streams, and sulfur-free biodiesel as solid acids do not leach into the product.
- Effective use of the integrated reactor volume leading to high unit productivity.
- Efficient use of the raw materials: complete conversion and high selectivity, stoichiometric reactants ratio, FFA conversion to esters and not to soap waste.
- Reduced equipment costs, with up to ~40% savings on the total investment costs.
- Competitive operating costs due to the integrated design and the elimination of conventional steps: handling of homogeneous catalyst and corrosive solutions, separation and disposal of salts, waste water treatment, excess alcohol recovery.

References

1. M. Balat, H. Balata, 2008, Energy Conversion and Management, 49, 2727.
2. M.G. Kulkarni, A. K. Dalai, 2006, Industrial & Engineering Chemistry Research, 45, 2901.
3. K. Narasimharao, A. Lee, K. Wilson, 2007, J. Biobased Materials & Bioenergy, 1, 19.
4. T. Okuhara, 2002, Chemical Reviews, 102, 3641.
5. A. Kiss, A. C. Dimian, G. Rothenberg, 2006, Advanced Synthesis & Catalysis, 348, 75.
6. A. Kiss, G. Rothenberg, A. C. Dimian, F. Omota, 2006, Topics in Catalysis, 40, 141.
7. A. Kiss, A. C. Dimian, G. Rothenberg, 2008, Energy & Fuels, 22, 598.
8. S. Steinigeweg, J. Gmehling, 2003, Ind. Eng. Chem. Res., 42, 3612.
9. C. Dimian, F. Omota, A. Bliek, 2004, Chem. Eng. & Proc., 43, 411.

Optimal Scheduling Under Variable Electricity Pricing and Availability

Pedro M. Castro,[a,b] Ignacio E. Grossmann,[b] Iiro Harjunkoski[c]

[a]*DMS, Inst. Nacional Engenharia Tecnologia e Inovação, 1649-038 Lisboa, Portugal*
[b]*Dep. Chemical Engineering, Carnegie Mellon University, Pittsburgh PA 15213, USA*
[c]*ABB Corporate Research Center, Wallstadter Str. 59, 68526 Ladenburg, Germany*

Abstract

This work addresses the scheduling of continuous plants subject to energy constraints related to time-dependent electricity pricing and availability. A Resource-Task Network continuous-time formulation is presented that can address these issues together with multiple intermediate due dates. To the best of our knowledge, it is the first model of its type that is able to effectively incorporate time-variable utility profiles. The validity of the approach is illustrated through the solution of a few example problems. When compared to a simple manual scheduling procedure, significant electricity savings can be achieved by switching production from periods of high to low electricity cost.

Keywords: energy, continuous-time formulation, decision-making tool.

1. Introduction

Enterprises are currently under pressure to produce at the lowest possible cost within continuously changing economic constraints. To achieve this goal, they must actively look at the best operating practices and optimize these both globally and locally. Within this overall goal, scheduling plays an important part.

This work is motivated by a real industrial problem. It involves the final stage of a multiproduct plant with electricity intensive parallel equipment units, where scheduling involves deciding when each unit has to produce a certain product. The products are then sent to storage units, where they remain until dispatching takes place. Meeting customer demands on time is vital, and for this reason, in some plants, no other factors are taken into account besides trying to keep the storage units full in order to be able to fulfill the orders. Plant scheduling is difficult due to the following factors: large combinatorial size arising from the number of equipment units, products and storage units; various operating and contractual constraints; liberalized electricity market with nontransparent billing practices. Due to the inherent complexity, the operator scheduling choices may be far from the optimal ones.

The most challenging aspect of plant scheduling is undoubtedly the incorporation of energy constraints related to electricity pricing and availability. We consider the case where in the planning stage, contracts are agreed between the electricity supplier and the plant, which often specifies maximum levels of power usage. If electricity consumption exceeds this threshold, the plant incurs in stiff penalties, whereas underproduction costs the same as planned production. Electricity cost varies significantly throughout the day, and this must be taken into consideration within the modeling framework. Another important aspect concerns meeting the sales forecasts, which typically occur at the end of each day.
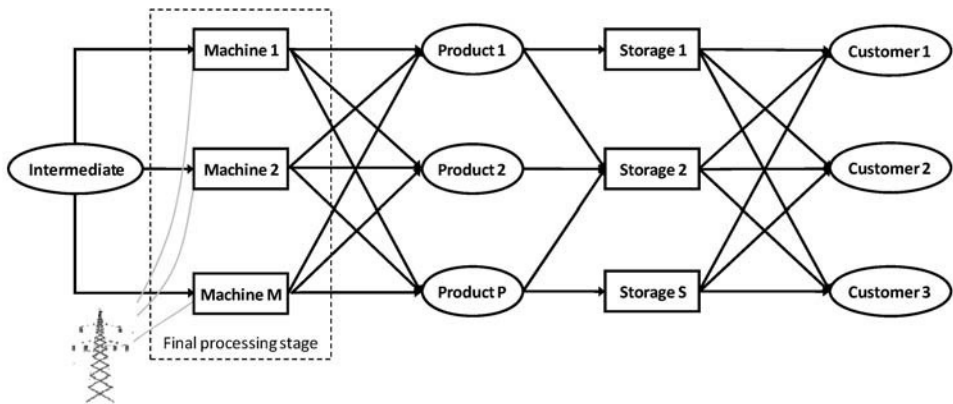
Figure 1. Final processing stage of industrial case study

Accounting for events that occur at predetermined points of time, being either a change in electricity cost level or the occurrence of a demand point, may be relatively easy or extremely complex, depending on the type of time representation employed. This is straightforward with a discrete-time approach. In contrast, with continuous-time, the timing of all event points is determined by the solver and thus it is much harder to relate the events with the points of change.

This paper presents a new continuous-time formulation that effectively handles time dependent cost parameters and discrete demand points. Incorporation of the former aspect within a continuous-time formulation has not been reported before to the best of our knowledge, whereas the constraints used to model the latter aspect are conceptually similar to those used by Maravelias and Grossmann (2003). The proposed approach builds on the general multipurpose formulation of Castro et al. (2004), which can address problems involving batch and continuous tasks, efficiently. Indeed, for batch plants, a study (2006) has found it to be the best single time grid formulation. Nevertheless, the constraints that are given are only suitable for continuous plants.

The new formulation is built on a unified framework for process representation, the Resource-Task Network (RTN) of Pantelides (1994). This means that the model variables and constraints are written is terms of abstract entities like resources, tasks and event points, and so it has a much wider scope than the single-stage industrial case study used for illustrative purposes.

2. Problem definition

During the final stage of the process, an intermediate material is transformed into one of different final products, characterized by chemical composition and particle size distribution, through the use of electricity. These are then sent to storage units, where they wait until customer dispatch takes place. This process is illustrated in Figure 1.

Typical plant schedules are established over one week, so this will be the time horizon (H=168 h) assumed. Let M represent the set of machines, P the set of products and S the storage units. Machines are characterized by: (i) power requirements $pw_{p,m}$ [MW]; (ii) processing rates, $\rho_{p,m}$ [ton/h]. Let DY and HR be the days of the week and the hours of the day, respectively. Each product may have multiple demands over the week, at any hour of the day, $d_{p,dy,hr}$ [ton]. Storage units have known maximum capacities, cap_s [ton]. The energy contract signed by the plant and electricity provider establishes a certain pricing policy. Electricity cost is typically lower during the night and higher during the day and we will be using the energy policy given in Duarte et al. (2009). It consists of

three energy levels, E, with prices c_e of 0.0481, 0.0945 and 0.2162 [€/kWh]. The maximum level of total power consumption is specified by parameter $pwx_{hr,dy}$ [MW].

3. Mathematical formulation

The proposed model uses the RTN framework to make it as general as possible. Before employing a RTN model it is required to describe the process and/or process recipe as a sequence of tasks involving certain resources. When this is done, one needs to determine the structural parameters, which pass the information about the structure to the mathematical model. Due to limited space we will be skipping this important modeling phase and moving directly to the mathematical formulation.

Tasks ($i \in I$) are characterized by two sets of extent variables, one binary, $N_{i,t}$, and one continuous, $\xi_{i,t}$. The former identify the start of task i at time point t, while the latter give the amount handled by the task. We are implicitly assuming that the tasks are either instantaneous (start and end at t) or continuous, which can be made to last a single time interval without loss of generality (Castro et al., 2004).

The continuous-time formulation relies on a single time grid to keep track of events taking place. It uses $|T|$ event points that can be placed anywhere between the origin and end (H) of the time horizon. Traditionally, tasks starting at event point t have been assumed to start at the absolute time determined for that event point, T_t. However, the tasks have total freedom to start anywhere provided that they end before T_{t+1}. The advantage of not requiring tasks to start (end) exactly at a time point is the reduction on the number of event points required to represent a solution, which seriously affects computational effort.

Consider a simple example involving the execution of a single processing task at slot/interval t, see Figure 2. Based on the problem data, one can build an algorithm to generate: (i) demand points ($td \in TD$) and respective timing tfx_{td}; (ii) time periods of constant cost for electricity level e ($tp \in TP_e$), and respective starting and ending times, $lb_{e,tp}$ and $ub_{e,tp}$ [h]. In cases where the duration of the task is lower than the duration of the lowest cost period located inside the interval, the interval boundaries shown, T_t and T_{t+1}, can result from two consecutive demand points. Variable Ts_t will then give the starting time of the task executed during interval t (or the earliest starting time amongst all tasks executed). Note that no further event points are required.

The fundamental timing constraint states that the difference in time between two consecutive event points must be greater than the duration of the task taking place, eq 1. It is written for all equipment resources involved in the timing constraints (R^{TC}) and implicitly assumes that there can only be one task executed in such equipment unit at a certain time. That is ensured by the initial resource availability and the excess resource balances, which are not shown here. The novel constraints are given next. Eq 2 is used to guarantee that the tasks are fully executed within time interval t, while eq 3 ensures that tasks start no sooner than the time of the interval's lower boundary.
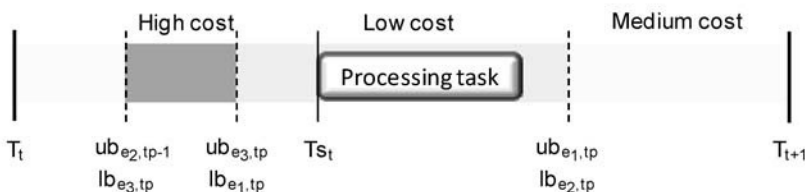
Figure 2. Continuous tasks executed at interval t do not necessarily start at event point t

$$T_{t+1} - T_t \geq \sum_{i \in I^c} \frac{\overline{\mu}_{r,i} \xi_{i,t}}{\rho_i^{\max}} \quad \forall r \in R^{TC}, t \in T, t \neq |T| \tag{1}$$

$$T_{t+1} - Ts_t \geq \sum_{i \in I^c} \frac{\overline{\mu}_{r,i} \xi_{i,t}}{\rho_i^{\max}} \quad \forall r \in R^{TC}, t \in T, t \neq |T| \tag{2}$$

$$Ts_t \geq T_t \quad \forall t \in T, t \neq |T| \tag{3}$$

We now define a new binary variable, $Y_{t,tp,e}$, which takes the value of one if, during interval t, tasks are processed within time period tp of energy level e. Similarly, binary variable $Y_{t,td}^{out}$, identifies whether or not event point t corresponds to demand point td. It can be assumed without loss of generality that there is: (i) one active time period of an energy level during t; (ii) one event point associated to due date td, eqs 4-5. If time interval t is located within time period tp of energy level e, then the starting time of tasks must be greater than the time period lower bound, eq 6. Likewise, they must end before its upper boundary (see Figure 2). Eq 7 is a big-M constraint that is only active if there is a continuous task being executed that belongs to energy level e. Finally, if event point t corresponds to demand point td, then the time values must match (eqs 8-9).

$$\sum_{e \in E} \sum_{tp \in TP_e} Y_{t,tp,e} = 1 \quad \forall t \in T, t \neq |T| \tag{4}$$

$$\sum_{t \in T \wedge t \neq 1} Y_{t,td}^{out} = 1 \quad \forall td \in TD \tag{5}$$

$$Ts_t \geq \sum_{e \in E} \sum_{tp \in TP_e} lb_{e,tp} Y_{t,tp,e} \quad \forall t \in T, t \neq |T| \tag{6}$$

$$Ts_t + \sum_{i \in I_e^c} \frac{\overline{\mu}_{r,i} \xi_{i,t}}{\rho_i^{\max}} \leq \sum_{tp \in TP_e} ub_{e,tp} Y_{t,tp,e} + H \cdot (1 - \sum_{i \in I_e^c} \overline{\mu}_{r,i} N_{i,t}) \quad \forall r \in R^{TC}, t \neq |T|, e \in E \tag{7}$$

$$T_t \geq \sum_{td \in TD} tfx_{td} Y_{t,td}^{out} \quad \forall t \in T, t \neq 1 \tag{8}$$

$$T_t \leq \sum_{td \in TD} tfx_{td} Y_{t,td}^{out} + H \cdot (1 - \sum_{td \in TD} Y_{t,td}^{out}) \quad \forall t \in T, t \neq 1 \tag{9}$$

The objective function is given in eq 10, which maximizes the total electricity cost [k€]. This is given by the sum over all electricity levels e, time intervals t and tasks i, of the product of electricity cost c_e [€/kWh], power consumption [MW], and task duration [h].

$$\min \sum_{e \in E} \sum_{i \in I_e^c} \sum_{r \in R^{UT}} \sum_{t \in T \wedge t \neq |T|} c_e \cdot (-\mu_{r,i}) \cdot \frac{\xi_{i,t}}{\rho_i^{\max}} \tag{10}$$

4. Computational Results

The performance of the model is illustrated through the solution of 5 test cases. Directly related to the complexity is the number of products, machines and storage units ($|P|$, $|M|$, $|S|$). Due to the differences in electricity cost among the energy levels, assigning production to the lower levels will have the biggest impact on the total cost. The resulting mixed-integer linear programming (MILP) problems were solved by CPLEX 11.1 up to a relative optimality tolerance=10^{-6} on an Intel Core2 Duo T9300 (2.5 GHz) laptop running Windows Vista Enterprise. The results are listed in Table 1.

The optimal solution for EX1 corresponds to a total electricity cost of €18,625. Eight out of 10 task instances are executed in low-cost periods with the remaining being in medium-cost levels. Enough power was provided to operate the machine throughout the time horizon. However, power shortages may force tasks to switch to a higher cost level. In order to show that the model can cope with constraints of this type, EX2 was solved. It is uses the same data of EX1 but now the maximum power consumption during the first seven hours of Tuesday and Thursday is set to zero. The cost increases by 15.8%, to €21,575, which is very significant considering such small changes in data. It serves to illustrate the impact that advanced scheduling tools can have on plant profitability.

Nevertheless, we do have to say that the continuous-time formulation is limited to small problems. Even for EX5, which features three products, two machines and two storage units, we had to specify low product demands (leading to a 33% capacity) to ensure that most tasks could fit into a single period of constant electricity cost. In this way, a good solution could still be found with a relatively small number of event points. However, it took already more than 1 hour to prove optimality for $|T|$=11 (€26,911), a value that is still 0.5% above the true optimum (found by a discrete-time formulation). The schedule and corresponding storage profiles are given in Figure 3. For larger problems, CT is intractable due to the following facts: (i) the computational effort is strongly dependent on $|T|$, with experience telling us that we get typically a one order of magnitude increase for a single increase in $|T|$; (ii) it is not straightforward to find a number that ensures feasibility; (iii) it may be even difficult for the solver to find out that the value of $|T|$ is insufficient, i.e. finding that the problem is infeasible.

Table 1. Computational results

problem	(P,M,S)	$|T|$	DV	SV	EQ	RMIP (€)	MIP (€)	CPU s	nodes
EX1	(2,1,1)	11	490	797	347	16620	**18625**	8.53	4252
		12	539	875	380	16620	18625	9.8	5442
EX2	(2,1,1)	11	490	797	347	16620	**21575**	26	20971
		12	539	875	380	16620	21575	21.6	15474
EX3	(2,1,2)	11	510	934	420	16620	**18153**	237	69046
		12	561	1025	460	16620	18153	1372	467962
EX4	(2,1,2)	11	510	934	420	18896	21349	178	46512
		12	561	1025	460	18896	**21175**	1470	237314
EX5	(3,2,2)	9	528	1089	562	25625	27222	7.18	3989
		10	594	1221	629	25625	27008	369	138426
		11	660	1353	696	25625	**26911**	4131	1295540

5. Conclusions

This paper has focused on the modeling of discrete events that occur at predetermined points in time with a continuous-time scheduling formulation. These included multiple intermediate due dates, utility availability and variable electricity costs. New sets of

constraints have been proposed that are part of a general model that relies on the Resource-Task Network process representation. The validity of the approach has been demonstrated on a few test cases adapted from a real industrial problem. Despite the major modeling breakthrough, the results have shown that only problems of small size can be handled effectively. More importantly, the paper has highlighted the importance of taking variable electricity costs into consideration when deriving the schedule. State-of-the-art scheduling formulations have the potential to achieve major savings when compared to procedures that are mostly focused on feasibility.

References

P. Castro, A. Barbosa-Póvoa, H. Matos, A. Novais, 2004, Ind. Eng. Chem. Res., 43, 105.

B. Duarte, L. Santos, J. Mariano, 2009, Computers Operations Research, 36, 1825-1834.

C. Maravelias, I. Grossmann, 2003, Proceedings ESCAPE-13 (Eds. A. Kraslawski and I. Turunen), pp. 215.

C. Pantelides, 1994, In Proceedings of the Second Conference on Foundations of Computer Aided Operations; Cache Publications: New York, 1994; pp 253.

M. Shaik, S. Janak, C. Floudas, 2006, Ind. Eng. Chem. Res, 45, 6190.
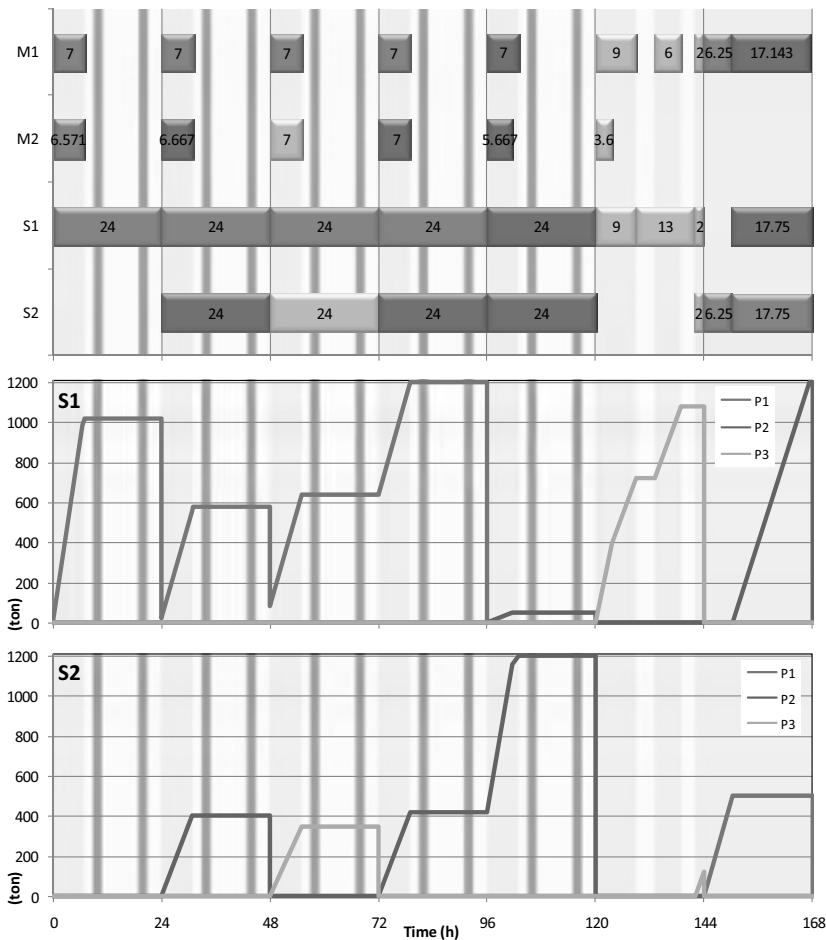
Figure 3. Best solution found for EX5.

1701

Effect of Substrate Specific Area on Lignocellulose Enzymatic Hydrolysis: An Experimental and Modeling Investigation

Chiara Piccolo[a], Gunnar Lidén[b], Fabrizio Bezzo[a]

[a] *DIPIC-Dipartimento di Principi e Impianti di Ingegneria Chimica, Università di Padova, via Marzolo 9, I-35131, Padova, Italy (chiara.piccolo@unipd.it)*
[b] *Department of Chemical Engineering II, Lund Institute of Technology, P.O. Box 124, 221 00 Lund, Sweden*

Abstract

Lignocellulose hydrolysis is a critical step in the enzymatic process for the production of bioethanol. Its high variability and dependence on materials and process parameters makes it very difficult to be described and optimized through a reliable modeling approach. In this work an experimental investigation has been carried out to assess the effect of the substrate specific area on the kinetics of lignocellulose hydrolysis. Data were used to estimate the parameters of a modified Langmuir adsorption model, embedding the accessible surface area as a critical parameter.

Keywords: surface area, adsorption experiments, Langmuir adsorption.

1. Introduction

Kinetics and further modeling studies of hydrolysis are useful in a different stages of processing of biomass to fermentable sugars. Mechanistic models are developed from the reaction mechanism, mass transfer consideration and other physical parameters that affect the extent of hydrolysis, and they vary in their complexity based on the intended use of the models [1].

Extensive research has been made to identify parameters that affect the rate of lignocellulose hydrolysis. Two main categories can be identified depending on whether parameters are mainly related to the substrate characteristics (lignin distribution, accessible surface area, particles size, crystallinity, degree of polymerization) or to the enzyme features (adsorption of enzyme prior to reaction, end-product inhibition, thermal and shear force inactivation, synergism, mass transfer limitation affecting the transport of the enzyme to the substrate) [2].

In the past 50 years, there has been a constant influx of research publications addressing the enzymatic kinetics of cellulose degradation. However, the kinetics of cellulose degradation is still not fully understood because of different competing effects that can hardly be distinguished from each other and that introduce large bias and variability in the estimation of kinetic parameters.

In the current work, an emphasis is given to the effect of the substrate specific surface area on the enzyme adsorption onto the substrate, which represents the essential prerequisite step of the hydrolysis process. In previous works it has been shown that the rate of adsorption is rapid compared to the actual hydrolytic activity of the enzymes,

thus making the amount of adsorbed cellulase an important factor in the effectiveness of the reaction [3]. The most common description of cellulase adsorption is the Langmuir isotherm (Eq. 1 in the following), derived assuming that adsorption may be described in terms of a single adsorption equilibrium constant and a specified adsorption capacity [4]. Embedding the effect of the specific surface in the model parameters accounting for the accessibility of the substrates represents a first attempt to deal with the analysis of adsorption in spatial terms so as to understand cellulose hydrolysis at a mechanistic level.

 A three-step approach was adopted to investigate the phenomenon:
  - surface area measurement and pores size distributions were determined for three different substrates through BET-technique;
  - adsorption isotherms and hydrolysis experiments were performed;
  - a modified Langmuir model was developed and adsorption equilibrium data and hydrolysis data were used to estimate the parameter of the model.

2. Materials and Methods

2.1. Enzymes

A commercial cellulase mixture (Celluclast 1.5L provided by Novozymes A/S, Bagsvaerd, Denmark) was used in the experiments. The cellulase enzyme had an activity of 46.8 FPU/ml.

2.2. Substrates

A comparative study was carried out on three different substrates: Avicel, and spruce pre-treated at two different conditions.
  - Avicel (i.e. microcrystalline cellulose) was purchased from Sigma-Aldrich.
  - Softwood hydrolyzates were prepared by steam explosion with SO_2 impregnation from spruce chips at Lund University. Two different treatment conditions were used: a) temperature 210°C, SO_2 content 2.5% (w/w) and residence time 5 min. The composition as determined by NREL standard procedure was 46.7% glucan, 1.9% mannan, 1.6% xylan, 1.2% galactan and 1.2% arabinan. Acid insoluble lignin was 42.0%; b) temperature 190°C, residence $time_e$ 10 min with the same SO_2 content as before. NREL composition analysis provided these results: 48.6% glucan, 2.9% mannan, 2.1% xylan, 1.2% galactan and 1.2% arabinan. Acid insoluble lignin was 35.7%.

The first pretreatment condition was known to produce hydrolyzate which more easily can be enzymatically degraded. Both pretreated spruce substrates were washed with distilled water and stored at 4°C.

2.3. BET measurements

This procedure, often used for the determination of the specific surface area of inorganic catalysts, is based on Nitrogen adsorption onto the material surface at different pressures.

Surface area, total pore volumes, at a relative pressure of about 0.994, and pore size distributions were determined on a Micrometrics ASAP 2400 after degassing for 10 days at 40°C.

Before degassing pretreated spruce materials were dried. Different drying conditions were chosen in order to investigate the impact on biomass inner structure of the dewatering process: samples were dried for 48 h in the oven at 105°C, and for 5 days at room temperature. Avicel, stored at room temperature, was straight subject to degassing.

2.4. Adsorption isotherms studies

Adsorption isotherms studies were conducted at constant reaction time and constant temperature with different enzyme-substrate ratios, for the three substrates whose BET surface had been assessed (i.e. Avicel and spruce pretreated at optimal and milder conditions). Different enzyme solutions were added with a constant amount of substrate in 50 mL plastic Falcon test tubes, to obtain a dry matter loading at 10g/L. The pH was adjusted at 4.8. After 90 minutes of incubation (60-90 minutes is the time needed to achieve maximal adsorption at the given conditions) at the selected temperature, the supernatant was collected, centrifuged and filtered. Cellulase residual activity in solution was assessed using the Bradford method for protein content determination. The experiments were performed first at 4°C to avoid changes in substrate properties due to extensive hydrolysis. The amount of bound cellulase is calculated from the difference between the initial cellulase or protein concentration and final free cellulase or protein concentration

2.5. Hydrolysis studies

Hydrolysis experiments were performed in 300ml baffled Erlenmeyer flasks, at pH 4.8 and 30°C, in distilled deionized water, using Avicel, the optimal pretreated spruce, and the non optimal pretreated spruce as substrate. The WIS (Water Insoluble Solids) content was 10g/L (i.e. about 1%w/v).

Two different enzyme loads were tested: 2ml and 0.067ml, corresponding roughly to 46.84 FPU/g WIS and 1.57 FPU/g WIS. Samples were withdrawn during the first 30 hours, centrifuged and the concentration of hydrolysis products was measured through HPLC.

3. Modeling approach

The most common description of cellulase adsorption is the Langmuir isotherm, derived assuming that adsorption can be described by a single adsorption equilibrium constant and a specified adsorption capacity [4]. The Langmuir isotherm may be represented as:

$$E_a = \frac{W_{max} K_P E_f}{1 + K_P E_f} \tag{1}$$

where E_a is the adsorbed cellulase (mg cellulase/L); W_{max} is the maximum cellulase adsorption $= A_{max}*C$ (mg cellulase/L) with A_{max} the maximum cellulase adsorption per g cellulose (mg cellulase / g cellulose) and C the cellulose concentration (g cellulose/L); E_f is the free cellulase (mg cellulase/L); and K_P is the dissociation constant ($K_P = \dfrac{E_a}{E_f C}$) in terms of L/g cellulose.

The focus of the current study is not to propose a new phenomenological adsorption model but rather to present a models which is as simple as possible with physically meaningful parameters Although very complex and detailed models could be envisaged, the real issue is to produce reliable experimental data that can be used to properly identify them.

Accordingly, the structure of the Langmuir equation, which is widely used because it provides a good (or often very good) fit to experimental data and represents a simple mechanistic model that can be used to compare kinetic properties of various cellulase–cellulose systems, was maintained. Specific surface areas and a measure of pore accessibility were embedded in the model. Variables and parameters on Eq. 1 are thus

changed as follows: E_a is the adsorbed cellulase (mg cellulase/g WIS); W_{max} is the maximum cellulase adsorption (mg cellulase/g WIS) defined as:

$$W_{max} = A_{max} S_a \alpha \tag{2}$$

with A_{max} the maximum cellulase adsorption/m^2 on the substrate (mg cellulase / m^2 substrate), S_a the substrate specific area (m^2/g WIS) and α is the fraction of pores accessible to enzyme as calculated from pores size distribution; E_f is the free cellulase

(mg cellulase/L); and K_P is the dissociation constant $(K_P = \dfrac{E_a}{E_f})$.

3.1. Estimation of the Model Parameters

A least squares regression was used to estimate the model parameters from the adsorption and hydrolysis experimental. In particular, in order to fit the glucose profiles, the kinetics proposed by Philippidis et al. [5, 6] was considered and adjusted to the new model variables definition. According to the Philippidis' model, based on the principles of heterogeneous catalysis, the reaction rate depends on the concentration of adsorbed cellulase (E_a) through:

$$r = k^1 E_a \tag{3}$$

where k^1 is a lumped specific rate constant [h^{-1}], which exhibits a Michaelis-Menten dependence on the cellulase concentration according to the following expression:

$$k^1 = \frac{kE}{K_{eq} + E} \tag{4}$$

where k is the maximum specific cellulose hydrolysis rate [h^{-1}], E is the total (free and bound) concentration of the enzyme and K_{eq} is the cellulase enzyme saturation constant [FPU/L]. The values of k and K_{eq} used for the estimation of the adsorption isotherm parameters were taken from Philippidis et al. [6].

4. Results and Discussions

4.1. BET measurements

Due to variations in the experimental conditions such as adsorption time, vacuum time and vacuum pressure, sample preparation and sample origin and features, a wide range of gross area values have been reported in the literature even for the same substrate. Typical superficial surface areas for Avicel exhibit a wide variability (in the range 1.8-22 m^2/g) [2]. For the steam pretreated spruce material, however, no previous measurements were found in literature. The surface area values, determined through BET procedure, are reported in Table 1.

Effect of Substrate Specific Area on Lignocellulose Enzymatic Hydrolysis:
An Experimental and Modeling Investigation

1705

Table 1. BET specific surface area values and fraction of the accessible surface determined for the substrates of interest.

Substrate	Drying at room T [sq. m/g]	Drying in the oven [sq. m/g]	Average value used in model identification	Average fraction accessible surface area
Avicel	1.1±0.0		1.1	0.453
Opt. pret. spruce	2.0±0.1	2.4±0.1	2.2	0.680
Non optimal pret. spruce	1.6±0.1	1.2±0.0	1.4	0.660

*Error of measurement ± 0.9%

For the "optimally" pretreated spruce, the BET surface areas were determined both for material dried at room temperature and material dried in the oven (105 °C). The values were only slightly different, i.e. the drying temperature does not have big impact on substrate structure, at least with concern to the surface area. The analysis of the pore size distributions, measured through the same procedure, is important in determining the fraction of pores accessible to the enzymes. Only pores with a diameter larger than 54Å are accessible to cellulose enzymes [7]. The fraction of accessible area, calculated from PSD data are also reported in Table 1.

4.2. Adsorption and hydrolysis experiments results

After incubation of various amounts of cellulase enzymes with a fixed amount of substrate, the free protein concentration was determined. From these data the adsorption isotherms (ratio amount of enzyme adsorbed/amount of enzyme added) were constructed **for** the different substrates (Fig. 1a). The error bars show the standard deviation of the data calculated from repeated experiments and analysis. By comparing the isotherms it can be observed that at the same WIS content the extent of adsorption is higher for the substrate characterized by the bigger BET surface area, i.e. the optimal pretreated spruce. The lowest amount of enzyme adsorbed was observed for Avicel, i.e. the substrate with the smallest surface area.

The profiles of glucose released during hydrolysis experiments at different enzyme loads (E1=46.84 FPU/g WIS and E2=1.57 FPU/g WIS) are shown in Fig. 1b.

(a)   (b)
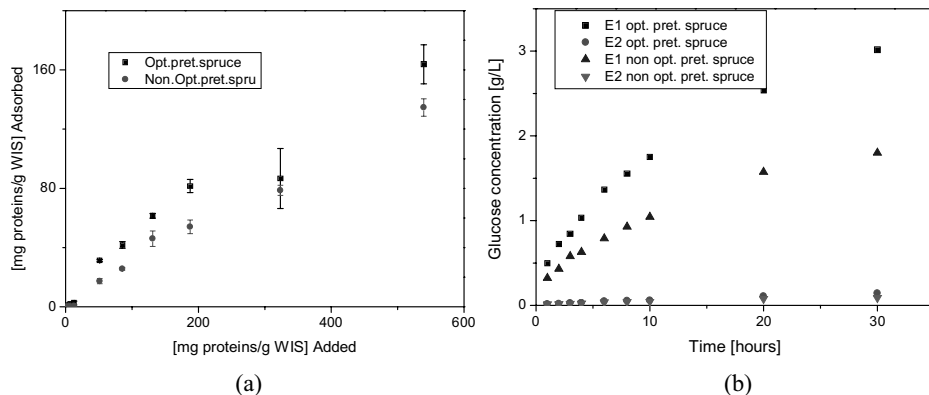
Figure 1 Adsorption isotherms (a) and glucose rate of release (b) for hydrolysis of pretreated spruce substrates.

4.3. Model identification

Data on free and bound enzyme concentration and glucose concentration were used to estimate the equilibrium adsorption constants, K_p and A_{max}. The values for the parameters are given in Table 2

Table 2. Parameter values for adsorption on pretreated spruce.

Parameter	Estimate	t-value	t_{ref}	Confidence intervals 95%	χ^2-value	χ^2_{ref}
A_{max}	92.208	7.511	1.653	12.204	229.66	216.53
K_p	0.0065	6.890		0.0009		

5. Final remarks

The investigation has highlighted the importance of the substrate specific surface area in the hydrolysis process for lignocellulosic materials. The experimental data were used to estimate the parameters of a modified Langmuir-type adsorption model taking into account the effect of the specific surface. Future work will aim at incorporating within a modeling framework the extent of productive and unproductive binding of enzyme, which can be assessed through a combined analysis of adsorption data and dynamic hydrolysis experiment data, performed on substrates characterized by sensibly different lignin content.

6. Acknowledgement

References

[1] S. Peri, S. Karra, Y.Y. Lee, M.N. Karim, 2007, Modeling intrinsic kinetic of enzymatic cellulose hydrolysis, Biotechnol. Progr., 23, 626-637.
[2] Y.H.P. Zhang, L.R. Lynd, 2004, Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems, Biotechnol. Bioeng., 88, 797-824.
[3] W. Steiner, W. Sattler, H. Esterbauer, 1988, Adsorption of *Trichoderma reesei* cellulase on cellulose-experimental-data and their analysis by different equations, Biotechnol. Bioeng., 32, 853–865.
[4] B. Nidetzky, W. Steiner, M. Claeyssens, 1994, Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: adsorptions of two cellobiohydrolases, two endocellulases and their core proteins on filter paper and their relation to hydrolysis, Biochem. J., 303, 817-823.
[5] G.P. Philippidis, D.D. Spindler, C.E. Wyman, 1992, Mathematical modeling of cellulose conversion to ethanol by simultaneous saccharification and fermentation process, Appl. Biochem. Biotechnol., 34-35, 543-556.
[6] G.P. Philippidis, C. Hatzis, 1997, Biochemical engineering analysis of critical process factors in the biomass-to-ethanol technology, Biotechnol. Progr., 13, 222-231.
[7] H.E. Grethlein, 1985, The effect of pore size distribution on the rate of enzymatic hydrolysis of cellulose substrates, Bio/ Technol., 3, 155–160.

An Outer Approximation Algorithm for the Global Optimization of Regulated Metabolic Systems

Gonzalo Guillén-Gosálbez[a*], Carlos Pozo[a], Laureano Jiménez[a], Albert Sorribas[b]

[a] *Department of Chemical Engineering, University Rovira i Virgili, Tarragona, Spain*
[b] *Departament de Ciències Mèdiques Bàsiques, Universitat de Lleida, Lleida, Spain*

Abstract

Understanding the evolution of cellular metabolism requires a number of techniques able to deal with its complexity. Adaptive responses observed in evolutive studies are expected to consist of an optimal set of changes in enzymes activities fulfilling important physiological constraints. Within this context, we present a novel approach to identify enzyme activity regions that contain feasible biological responses in evolution. The framework presented also allows to optimize the enzyme activity changes required to maximize certain fluxes in biotechnological applications. The method relies on solving nonlinear programming models via global optimization techniques.

Keywords: Optimization, Power-law, Evolution.

1. Introduction

In natural cells, emergence of new designs results from evolution. The adaptive response of the cellular metabolism to different situations is attained by tuning gene expression and enzyme activity. Understanding the evolution of adaptive strategies is an important goal in Systems Biology.

The evolution of adaptive stress responses can be seen as a multi objective optimization problem. In that sense, the observed response represents an optimal (in some sense) combination of changes that ensure appropriate survival in the considered conditions. Evolution results in adaptations that are admissible solutions fulfilling important physiological constraints.

Within this general context, we introduce a novel approach that aims to identify enzyme activity regions containing feasible responses observed in evolution. The method introduced can also be employed to optimize biological systems in biotechnological applications. Our approach focuses on the properties of a particular class of non-linear models, the GMA (Generalized Mass Action) models that are based on the power-law formalism. The proposed algorithm is very efficient for realistic problems. The solutions found would represent the landscape in which evolutive solutions are expected. Comparison of our results and actual data shows the practical usefulness of the proposed method.

2. GMA representation

We shall consider a metabolic network that has p fluxes that can contribute to the change in the concentration of the pool of any of the n internal metabolites:

$$\frac{dX_i}{dt} = \sum_{r=1}^{p} \mu_{ir} v_r \qquad\qquad i = 1,...n$$

Here, μ_{ir} is a stoichiometric factor that indicates how many molecules of X_i are produced or used by the process v_r; it is a positive integer if the flux r produces X_i and it is a negative integer if the flux r depletes the pool of X_i. Each velocity can be represented by different functional forms, but, the so-called power-law formalism is one of the most convenient:

$$v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}}$$

In this representation, X_j accounts for the concentration of metabolite j, γ_r is an apparent rate constant for flux r, and f_{rj} is the kinetic order of variable X_j in reaction r. Each kinetic order quantifies the effect of the metabolite X_j on flux r and corresponds to the local sensitivity of the rate v_r to X_j evaluated at the corresponding operating point. Using this representation, a Generalized Mass Action (GMA) model is defined as [1]:

$$\frac{dX_i}{dt} = \sum_{r=1}^{p} \mu_{ir} \left(\gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \right) \qquad\qquad i = 1,...n$$

In this expression, m indicates independent (external) metabolites.

3. Optimization model and solution strategy

Here, we present an optimization framework for GMA systems that will be later on taken as a basis for deriving the feasibility approach, which is the main contribution of this work. Non-linear optimization models based on the power-law formalism were first proposed by Voit [2]. In S-system representations, a transformation to logarithmic coordinates can be applied thus leading to linear optimization models. However, when the problem is represented by a GMA model, this technique cannot be applied.

In general, the problem of identifying the optimal values of v_r, γ_r and X_j that maximize a given criterion and satisfy at the same time the equations involved in the GMA representation can be posed as a nonlinear programming (NLP) as follows:

$$ONLP = \min \ U(v_r, \gamma_r, X_j)$$

$$s.t. \quad \sum_{r=1}^{p} \mu_{ir} v_r = 0 \qquad\qquad i = 1,...n$$

$$v_r = \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}} \qquad\qquad r = 1,...,p$$

$$v_r, \gamma_r, X_j \in \Re_+$$

Model **ONLP** corresponds to a non-convex problem. Because of this, standard NLP techniques may get trapped in local solutions that are likely to be far away from the global optimum. This may lead to wrong conclusions when performing biological studies. To circumvent this limitation, we introduce a deterministic algorithm to globally optimize **ONLP** that is based on the works of Bergamini and co-workers [3] and Polisetty et al. [4]. The proposed method relies on hierarchically decomposing the problem into two levels, an upper level master problem **CMILP** and a lower level slave problem **RNLP**, between which the algorithm iterates until a termination criterion is satisfied (see Figure 1).

The master level of the algorithm entails the solution of a mixed-integer linear (MILP) problem, which is a relaxation of model **ONLP** (i.e., it rigorously overestimates the feasible region of **ONLP**), and therefore predicts a valid lower bound on its global optimum. In the lower level, the original problem is locally optimized in a reduced search space (i.e., model **RNLP**), thus yielding an upper bound on its global solution. The upper and lower level problems are solved iteratively until the bounds converge. Due to space limitations, technical details of the main features of the proposed algorithm are omitted. As mentioned before, this method can be employed in biotechnological applications in order to optimize a given bioprocess. In this work, as discussed in section 4, such method is employed to derive a tool to perform feasibility analysis in evolutive studies.
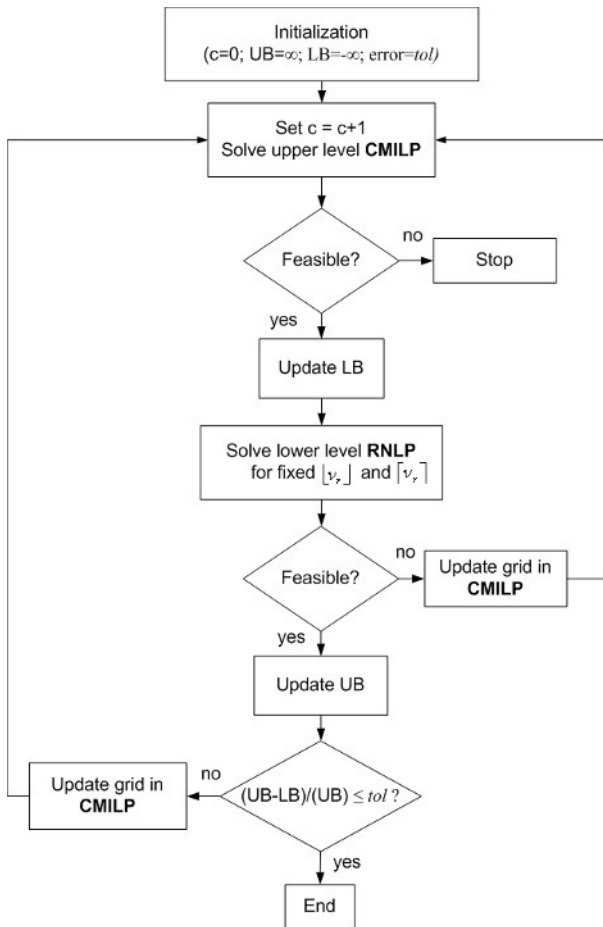
Figure 1. Proposed algorithm.

4. Feasibility approach

The algorithm previously presented can be used, after minor modifications, to identify regions that contain feasible solutions to the original problem **ONLP**, and discard others

in which no single feasible solution exists. Given the metabolic model, the goal is then to find the admissible changes at the level of enzyme activities that are compatible with a set of physiological and functional effective criteria.

From the mathematical point of view, this analysis requires the definition of a set of disjoint sets P_S^q ($P_S^q \cap P_S^{q'} = 0$ for all $q \neq q'$) such that their union contains the feasible space S of **ONLP** ($S \subseteq \bigcup_{q=1,\dots,Q} P_S^q$). In this work, for the sake of simplicity, we assume that each of these regions P_S^q is a hyper-rectangle described by a set of linear inequalities that impose lower and upper limits ($\lfloor \gamma^q \rfloor$ and $\lceil \gamma^q \rceil$, respectively) on the values of the apparent rate constants γ^q. Thus, we have:

$$P_S^q = \left\{ (v, \gamma, X) \in \Re_+^p \times \Re_+^p \times \Re_+^{n+m} : \lfloor \gamma^q \rfloor \leq \gamma^q \leq \lceil \gamma^q \rceil \right\} \qquad q = 1,\dots,Q$$

Hence, the feasibility analysis must determine whether these hyper-rectangles contain feasible solutions to **ONLP** or not.

The method devised to accomplish this task is based on the same ideas presented before and comprises two different levels. At the upper level, a master problem is solved to identify a region (i.e., hyper-rectangle) that may contain a feasible solution of **ONLP**. At the lower level, the prediction made by the master problem is checked by solving the original problem in a reduced search space. If a feasible solution is found, then integer cuts are added to the master problem in order to exclude the region containing such a feasible point. Otherwise, the master model is updated by refining its grid, until either a feasible solution is obtained in the lower level or the higher level problem turns out to be unfeasible.
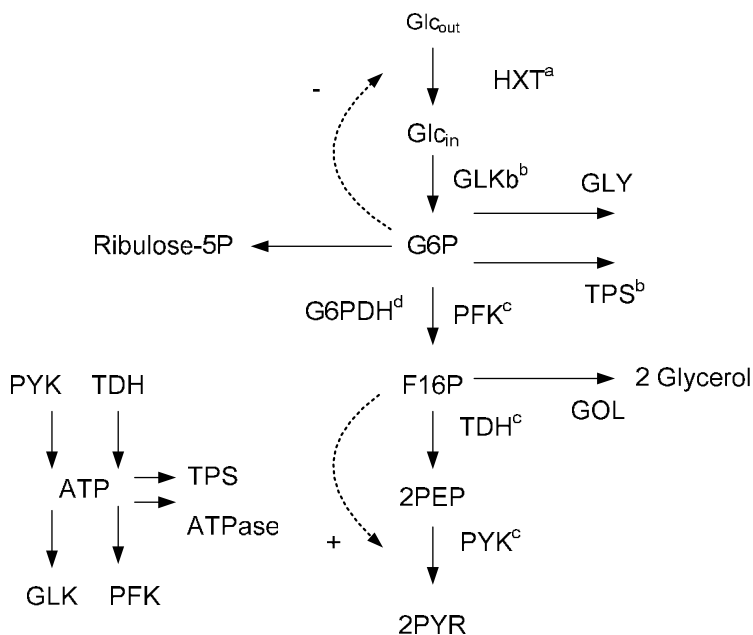
Figure 2. Scheme of the modeled pathways and ranges used for generation of the in silico gen expression profiles (GEPs).

5. Feasible adaptive response of yeast to heat shock

The capabilities of our method were illustrated through its application to the optimal adaptive response of yeast to heat shock (for a detailed description see [5]). Our model includes the core of the glycolytic pathway and the first step of the pentose phosphate pathway. It also accounts for the synthesis of glycogen, trehalose and glycerol, as shown in Figure 2. The notation used in this figure is as follows. Glcout: Extracellular Glucose; Glcin: Intracellular Glucose; G6P: Glucose-6-phosphate; F16P: Fructose-1,6-biphosphate; PEP: Phosphoenolpyruvate; PYR: Pyruvate; HXT: Hexose transporters (HXT1–4, HXT6–8, HXT12); GLK: Glucokinase/Hexokinase (GLK1, HXK1, HXK2); PFK: Phosphofructokinase (PFK1, PFK2); TDH: Glyceraldehyde-3-phosphate dehydrogenase (TDH1, TDH2, TDH3); PYK: Pyruvate kynase (PYK1, PYK2); GLY: Production glycogen; TPS: Trehalose 6-phosphate syntase complex (TPS1, TPS2, TPS3); G6PDH: Glucose 6-phosphate dehydrogenase (ZWF1).

The metabolic network was found to be specially sensitive to changes in two specific enzymes (i.e., PFK and TDH). For this reason, the feasibility analysis was performed on their domain, defining ten different sub-intervals for each of them. Hence, in this particular example, the feasibility analysis focuses on identifying, from the initial set of 100 hyper-rectangles, those containing feasible solutions to **ONLP** and those in which no feasible point exists.
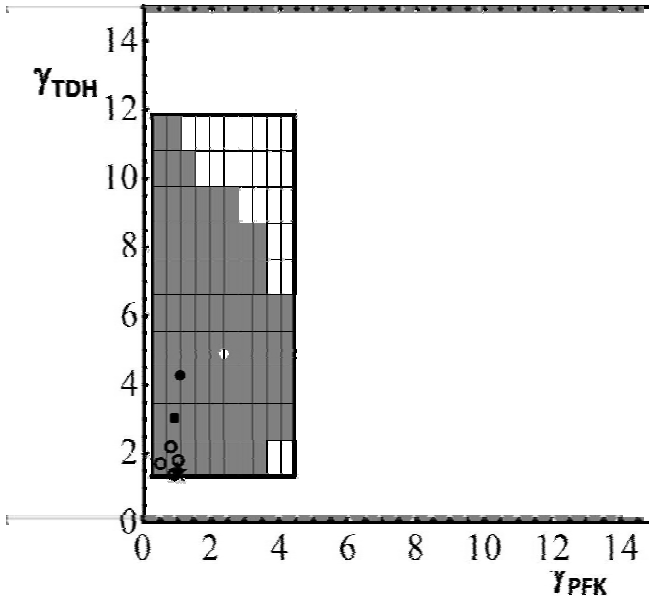
Figure 3. Feasibility analysis. White circle: Maximum rate of ATP synthesis; Black circle: Maximum rate of NADPH synthesis; Black square: Maximum rate of Trehalose synthesis; Black star: Minimum cost; Black empty circles: Experimental observations by Vilaprinyo et al. [5]

The algorithm was implemented in GAMS interfacing with CPLEX and CONOPT as main optimization packages. The total CPU time was less than one minute on an Intel 1.2 GHz machine. Results of this analysis are depicted in Figure 3.

In the figure, shady boxes represent hyper-rectangles that contain at least one feasible solution to the problem, whereas those in white have been proved to be unfeasible. For

comparison purposes, we have also depicted other solutions that are optimal in terms of some criteria: maximum rate of ATP synthesis, maximum rate of NADPH synthesis, maximum rate of Trehalose synthesis and minimum cost. This last metric (i.e., the cost) measures the overexpression of the enzymes.

As can be seen, experimentally observed responses [5] fall within the feasible region predicted by the algorithm. Interestingly, they allocate especially close to the minimum cost solution, that is, the one that would minimize the overexpression of the enzymes. Additionally, the maximum Trehalose rate solution is also near. This probably indicates some importance in the adaptation process.

6. Conclusions

This work introduced a systematic method for identifying the enzyme activity changes that allow a system to meet a set of physiological constraints while optimizing a parameter in the network. The approach presented relies on formulating nonconvex nonlinear problems that are solved via global optimization techniques.

The approach presented was applied to study the optimal adaptive response of yeast to heat shock. Experimental data fall well within the feasible region predicted. The closeness of those points to the minimum cost solution suggests that a conservative strategy where minimum changes are done is the preferred adaptive response. On the computational side, our method proved to be very efficient for medium size problems. The solutions found are intended to shed light on both, biotechnological and evolution studies.

7. Acknowledgments

References

Voit EO: Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists. Cambridge, U.K.: Cambridge University Press 2000.

Voit EO: Optimization in integrated biochemical systems. Biotechnol Bioeng 1992, 40(5):572-582.

Bergamini ML, Aguirre P, Grossmann IE: Logic-based outer approximation for globally optimal synthesis of process networks. Computers and Chemical Engineering 2005, 29:1914-1933.

Polisetty PK, Gatzke EP, Voit EO: Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. Biotech Bioeng 2008, 99(5):1154-1169.

Vilaprinyo E, Alves R, Sorribas A: Use of physiological constraints to identify quantitative design principles for gene expression in yeast adaptation to heat shock. BMC Bioinformatics 2006, 7:184.

1713

Controlled Formation of Self-assembled Nanostructures with Desired Geometries: Robust Dynamic Paths to Robust Desired Structures

Earl O. P. Solis, Paul I. Barton, George Stephanopoulos

Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

Abstract

This paper discusses the design principles underlying the controlled formation of nanostructures with desired geometries through a hybrid top-down and bottom-up approach: top-down formation of the physical domains with externally-imposed controls and bottom-up self-assembly of the nanoscale particles to form the desired structure. We propose a two-phase approach for this design problem. The first phase guarantees a robust desired structure, and the second allows the desired structure to be reachable from any initial particle distribution in the physical domain. Both phases require the solution of combinatorially-constrained quadratic optimization problems. The dynamics of the self-assembly process is described through a multiresolution view of the system. Crucial to the achievement of the design goals is the need to break the ergodicity of the system.

Keywords: Robust nanostructures, self-assembly dynamics, ergodicity breaking.

1. Introduction

The fabrication of structures with geometric features at the nanometer length scale is essential to the manufacturing of future electronic, magnetic and optical devices composed of nanoparticles. One major challenge for the formation of complex nanoscale structures is the precise positioning of particles into the desired geometries. The work described below focuses on the design principles and methodologies underlying the controlled formation of such structures through a hybrid top-down and bottom-up approach: top-down formation of physical domains with externally-imposed controls (degrees of freedom), and bottom-up generation of the desired structure through the self-assembly of the nanoscale particles, driven by interparticle interactions and interactions with the external controls. This type of approach is seen experimentally in templated self-assembly techniques (reviewed in Koh, 2007), e.g., crystallization on template surfaces (Aizenberg et al., 1999), crystallization of colloids in optical fields (Burns et al., 1990), and DNA-programmed placement using DNA crystals as scaffolds (Le et al., 2004).

Many of the current self-assembly techniques are only applicable to periodic, close-packed structures, which simply require judicious nanoparticle design. The hybrid design strategy we propose in this paper can be used to systematically engineer more complex non-periodic, non-close-packed structures, which require external controls to guide the self-assembly process. To implement the design strategy we need to solve the following key problems: (1) *Static Problem*: the placement of externally-imposed controls and determination of their intensities in order to ensure a stable self-assembled

desired structure; (2) *Dynamic Problem*: the use of time-varying external controls to ensure that the desired final structure can be reached with the highest probability from any initial particle distribution. This paper is organized as follows. Section 2 introduces the two case studies discussed throughout this paper. Section 3 and 4 will focus on the static and dynamic problems, respectively.

2. Self-assembly Model Systems

The model system we use to demonstrate our system design strategies is an isomorph of the Ising model. Figure 1a shows 1- and 2-D desired structures in finite lattice domains. The 1D example has a volume, $V=8$, and particle number, $N=5$; the 2D example has a volume, $V=64$, and particle number, $N=19$. The former system only has 56 total configurations in its phase space and is used to compare dynamic Monte Carlo (MC) simulation results to Boltzmann probability distribution calculations in the Canonical prescription. All dynamic MC simulations were performed using the "virtual-move" MC algorithm (Whitelam and Geissler, 2007).

Ergodicity-breaking is a key component of our design strategy. An ergodic system is one that exhibits eventual access to all other phase space states from any particular system state. For the controlled self-assembly processes described in this paper, the externally-imposed controls offer degrees of freedom, which are used to decrease the volume of phase space accessible to the system, i.e., decrease the number of states accessible from a given state. Such systems exhibit *non-ergodicity*, and their phase space can be decomposed into ergodic subsets called *components.* Transitioning between any two components separated by large energetic barriers is not very probable, as it would require either an unreasonably high system temperature or long period of time.
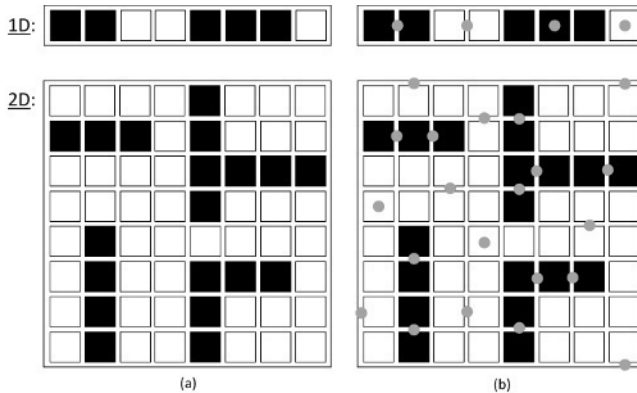
Figure 1: (a) Example desired configurations in finite 1- and 2-D lattice volumes. (b) The location of point conditions as external controls in the 1- and 2-D example systems.

For both example systems, we assume that the particles do not rotate but simply translate throughout the discrete system volume. We also assume that the particles are indistinguishable and negatively charged with a charge equal to -1. The binary interaction potential energy between particles is given by

$$\beta E = z_i z_j J_{ij} = z_i z_j \left[\frac{1}{r_{ij}} - \frac{1}{r_{ij}^6} \right], \tag{1}$$

where r_{ij} is the positional distance between lattice sites i and j. Variables z_i and z_j are binary with value 0 for an empty lattice site or 1 for an occupied site. This is a simple

model that is intended to simulate a long-range Coulombic repulsion and short-range attractive potential. For real systems, more complex interaction energy models will be required, but the above model serves as a phenomenological one.

The static design strategy involves two steps that create an ergodic component with the desired configuration as its only member: (1) define the locations of the system degrees of freedom needed to guarantee a robust desired configuration, and (2) optimize the intensities of each degree of freedom. The dynamic strategy is similar in that it restricts the system phase space, but it does so progressively over time until it reaches the component defined by the static problem.

3. Static Problem Strategy

Assuming that the system has achieved the desired structure, we must guarantee that this structure is maintained, i.e., we have to ensure that the desired structure is robust. To do this, we specify the necessary system degrees of freedom: the location of point conditions (attractive or repulsive point charges) and their intensities (charge values). The degrees of freedom we use are isotropic and well- or barrier-forming, depending on the charge. Because the presence of two wells forms a barrier and vice versa, only one type of point condition is needed to define the general features of the energy landscape.

In defining the point condition locations, we essentially separate the system into subvolumes of occupied lattice sites that require attractive point conditions and unoccupied sites that require repulsive point conditions. To find the locations of the minimum number of well- and barrier- forming point conditions needed, we developed an algorithm which covers the system volume with *tiles* of specific shapes. Specifically, for each point condition we assign a tile which encompasses the local area that the point condition is intended to influence. Thus, the minimum number of tiles needed to cover the barrier-forming regions determines the minimum number of negative point charges, and the minimum number of tiles needed to cover the well-forming regions determines the minimum number of positive point charges. We select the type of point condition with the smallest number necessary to cover the appropriate regions. The details of the set-covering method can be found in Solis et al. (2009).

For the 1- and 2-D example systems, Figure 1b shows the solution to the set-covering method. The 1D example has 2 well-forming and 2 barrier-forming point conditions. We choose to use the well-forming type. The 2D example has 11 well-forming point conditions and 10 barrier-forming point conditions. We choose the 10 barrier-forming point conditions since it is smaller in number.

Given the minimum number of point conditions and their locations, we want to generate an energy landscape in the physical domain that ensures a robust desired structure. The solution to the following problem determines the robustness of the desired structure against statistical fluctuations:

$$\max_{s \in S, \delta} \delta$$
$$\text{s.t. } E(s, z) - E(s, z_d) \geq \delta, \forall z \in \zeta^{\alpha} \backslash z_d, \tag{2}$$

where z_d represents the desired structure, s represents the point condition strengths, and ζ^{α} represents the subset of competing configurations that make up component α. The solution to this optimization problem ensures that the difference in potential energy, δ, between any configuration in the ergodic component, to which the desired structure belongs, and the desired structure is as large as possible. The δ value is therefore a

measure of the desired structure's robustness. If δ is sufficiently large, then the static problem is solved. If not, additional point conditions are needed. Since the minimum number of point conditions is of the same type, i.e., well- or barrier-forming, the additional point conditions will be selected from the opposite type.

For the 1- and 2-D examples, the solution to the optimization problem (2) gives the output $\delta = 19.8$ and $\delta = 0.1$, respectively. If, for instance, we would like the value of δ for the 2D example to be larger to satisfy a system robustness criteria, we can find the competing configuration(s) whose energy differs the least from the desired structure, i.e., the energy difference is equal to δ. The configurational differences between the desired and competing configuration(s) inform us of the *constraining feature(s)* of the desired structure. We can select the point condition(s) of the opposite type that energetically influences the constraining feature(s) and solve the optimization problem (2) again with the added point condition(s). In the 2D example, the constraining feature is the particle located in the lattice row 6 and column 7. Hence, we activate the point condition near it (row 6 and column 6.5) and solve the optimization problem again to find an increased value for the minimum energy difference, $\delta = 0.6$.

4. Dynamic Problem Strategy

Simply guaranteeing that the desired configuration is the minimum energy configuration within its component does not mean that the system, starting from any initial state, has a high probability of reaching the desired state at a particular temperature or within a reasonable timeframe. In fact, the formation of large barriers between the desired configuration and neighboring states in the solution to the static problem above creates a rough potential energy landscape, and given that the system has an unknown initial configuration, the problem of becoming kinetically trapped in a metastable configuration needs to be addressed.

We propose the use of time-varying degrees of freedom to restrict the volume of accessible system phase space states. To do this, we have adopted a multiresolution view of the desired structure. Figure 2 shows a systematic course-graining of the system particle number for the 1D example system. The proper use of time-varying degrees of freedom helps force the system to go through progressively smaller subsets of phase space, guaranteeing a higher likelihood of achieving the desired state when compared to simply using the static problem solution alone.

Similar to the static problem formulation above, the following optimization problem is solved for each stage in the dynamic process:

$$\max_{s \in S, \delta^{(i)}} \delta^{(i)}$$
$$s.t. \ E\left(s, z^{(i)}\right) - E\left(s, z_d^{(i)}\right) \geq \delta, \forall z^{(i)} \in \zeta \backslash z_d^{(i)}. \tag{3}$$

where $z_d^{(i)}$ represents all configurations that belong to the desired coarse-grained density specification of stage i and ζ represents all configurations in phase space. The main principle behind this optimization formulation is to maximize the energy gap, $\delta^{(i)}$, between the configuration in $z_d^{(i)}$ with the highest energy and the configuration in the remainder of phase space (excluding the configurations that satisfy the desired density specification) with the lowest energy. In turn, this will also try to push the overall minimum energy state to be within $z_d^{(i)}$.
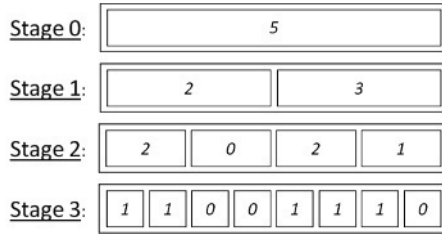
Figure 2: The systematic coarse-graining of the desired configuration, where Stage 3 is the desired detailed structure. The coarse-grained particle number specifications defined in the lower stages contain the desired structure but also contain progressively larger subsets of the phase space. Stage 0 contains all 56 states.

Given that the point condition locations in solving the static problem were sufficient to guarantee a robust desired configuration, we utilize the same set of point conditions for solving the dynamic problem. The solutions to this problem for Stages 1-3 of the 1D example system are shown in Figure 3.

The dynamic process for the 1D example system proceeds as follows: The system starts in Stage 0 with no point conditions and $k_B T = 1$; dynamic MC simulation results show that the system is ergodic at this temperature. Upon reaching equilibrium, we then impose the Stage 1 point conditions on the system at the temperature that maximizes the probability of being in the Stage 1 desired state, found through dynamic MC simulations to be $k_B T = 0.1$. Upon reaching equilibrium with the Stage 1 point condition intensities, we then impose the Stage 2 point condition intensities at a temperature ($k_B T = 0.25$) that maximizes the probability of the Stage 2 desired state starting from any configuration that satisfies the previous level's density specification. The same step is performed for Stage 3, where dynamic MC simulation results allow us to maintain $k_B T = 0.25$. After the system equilibrates, the solution to the Static Problem is used to maintain the desired configuration.
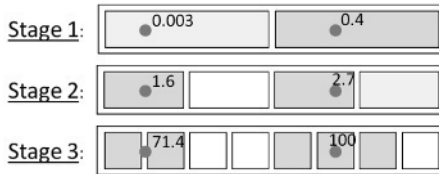
Figure 3: The dynamic point condition intensity solution to the optimization problem defined in Equation 3 for the 1D example system.

Figure 4 compares the results at the different stages of the simulation of the dynamic process to simulation results using the same point condition intensity values at each stage but starting from a random initial configuration. The time-varying process using the solutions to the optimization problems defined in Equation 3 shows a significant improvement in producing the desired configuration. However, there is room for further improvement, which comes from two further refinements: (1) A smoother transition between the point condition intensities and temperatures at each stage. Simulation results show that a simple linear transition is not favorable. This is because as the values of the external controls change in time, they induce higher energy barriers. Thus, the rate of change in the intensity values must decrease in order to allow equilibration in the current subset of phase space. (2) Though the system exhibits non-ergodic behavior at the temperatures used in the dynamic process, some of these temperatures are near the glass transition temperature, where there is a transition between ergodic and non-ergodic behavior. Utilizing a smoother evolution between the

stages of the dynamic process allows us to end each stage at lower temperatures where the system exhibits purely non-ergodic behavior, therefore increasing the probability of remaining in the desired subset of phase space. This gradual transition also allows systems "trapped" in metastable undesirable states to find the desirable state within a reasonable amount of time, thus increasing the probability of achieving the desired state.
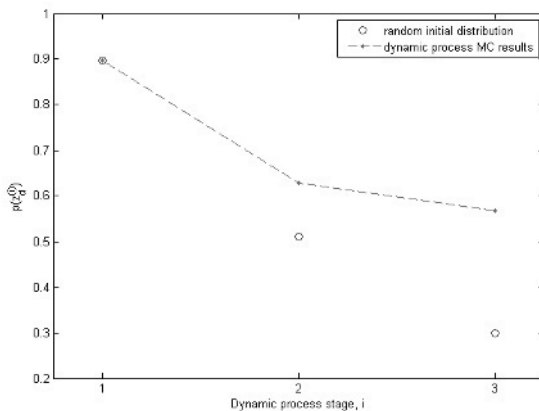
Figure 4: Dynamic MC results for the dynamic process compared to simulations with random initial states at the conditions of each stage in the dynamic process.

Conclusions

In this paper we proposed a two-phase approach for the controlled self-assembly of nanoparticles to form structures with desired geometries. The first phase provides us with the necessary system degrees of freedom to guarantee a robust desired structure. The second phase utilizes a dynamic process with time-varying degrees of freedom to increase the probability of reaching the desired structure from any initial particle distribution. Both phases require solving combinatorially-constrained quadratic optimization problems in order to find the external controls needed to dynamically reach the desired nanostructure and maintain it within a specific level of robustness.

Acknowledgements

References

J. Aizenberg, A.J. Black, G.M. Whitesides, 1999, Control of Crystal Nucleation by Patterned Self-Assembled Monolayers, Nature, 398, 495-498.

M.M. Burns, J.M. Fournier, J.A. Golovchenko, 1990, Optical Matter: Crystallization and Binding in Intense Optical Fields, Science, 249, 749-754.

S.J. Koh, 2007, Strategies for Controlled Placement of Nanoscale Building Blocks, Nanoscale Research Letters, 2, 519-545.

J.D. Le, Y. Pinto, N.C. Seeman, K. Musier-Forsyth, T.A. Taton, R.A. Kiehl, 2004, DNA-templated Self-assembly of Metallic Nanocomponent Arrays on a Surface, Nano Letters, 4, 2343-2347.

S. Whitelam, P.L. Geissler, 2007, Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles, Journal of Chemical Physics, 127, 154101.

E.O.P. Solis, P.I. Barton, G. Stephanopoulos, 2009, Controlled Formation of Nanostructures with Desired Geometries, FOCAPD '09 Proceedings.

Conversion of Glycerol to Liquid Fuels

Carlos A. Henao, Dante Simonetti, James A. Dumesic, and Christos T.
Maravelias

Department of Chemical and Biological Engineering University of Wisconsin
1415 Engineering Drive, Madison - Wisconsin, USA.
E-mail address: christos@engr.wisc.edu

Abstract

A significant fraction of the total petroleum supply is used for transportation in the form of liquid hydrocarbons. This fact, along with the increasing demand for oil in developing countries has led to substantial research efforts in the area of renewable liquid fuels. One alternative is the conversion of vegetal oil and animal fat into bio-diesel. However, this comes with the production of significant amounts of glycerol, a byproduct that will become abundant if large scale bio-diesel production is implemented. In order to increase the total biomass to fuel efficiency during the production of bio-diesel and address the overproduction of glycerol, a novel integrated Glycerol Reforming (GR) + Fischer-Tropsch (FT) process is presented. The novelty of this process lies in the use of a Rhenium-based catalyst for the conversion of aqueous glycerol to synthesis gas (syngas). This step reduces significantly the cost of syngas production in traditional green FT processes. This work presents a preliminary process synthesis and an economic evaluation for a medium capacity GR-FT plant. The results show that the integrated process is economically attractive, and that there is room for further improvements trough the use of systematic process design and optimization methodologies.

Keywords: Liquid Fuels, Renewable Energy, Glycerol Reforming, Fischer-Tropsch.

1. Introduction

Petroleum feedstocks are currently used to provide liquid hydrocarbons for transportation fuels, and concerns regarding diminishing petroleum reserves have fostered the development of renewable transportation fuels to supplement or replace those derived from petroleum (Corma et al. 2007, Huber et al. 2007, Bozell 2006, Chheda et al. 2007, Ragauskas et al. 2006, Rostrup 2004) Biomass provides an abundant, renewable source of carbon-containing molecules, thereby representing a potential alternative source of liquid transportation fuels (Klass 1988). The conversion of biomass-derived feedstocks to fuels in a manner that is cost-competitive with the refining of petroleum requires the development of new approaches that simplify the processing steps, thereby reducing capital and operating costs associated with separation and purification (Ragauskas et al. 2006). In this respect, the production of bio-diesel is an attractive option in that the processing steps are well established and production costs are low, compared to the case for production of liquid fuels from lignocellulose resources. For example, triglycerides can be converted to long-chain esters for diesel

fuel by transesterification reactions with methanol. However, this process leads to the formation of a glycerol co-product stream, representing an opportunity to convert a low-value waste stream to a more valuable product.

In the present paper, we explore the option of converting the waste glycerol stream from bio-diesel production to long-chain, linear hydrocarbons for use as liquid fuel. The approach we follow is to utilize the coupling of glycerol reforming to syngas over a PtRe-based catalyst, with syngas conversion to liquid alkanes by Fischer-Tropsch synthesis. This integrated process can potentially improve the economics of "green" Fischer-Tropsch process by reducing costs associated with syngas production, for example, by eliminating the need for an O_2-blown auto-thermal reformer or biomass. Also, our process presents the opportunity for reducing the size of the Fischer-Tropsch synthesis reactor by producing an undiluted synthesis gas stream and for eliminating subsequent cleaning steps required for synthesis gas produced from biomass gasification (Hamelinck et al. 2004, Bartholomew et al. 2006, Soares et al. 2006, Spath et al. 2003). Thus, our integrated process potentially allows for smaller scale Fischer-Tropsch synthesis plants to produce liquid fuels from biomass, as would be required for utilization of the glycerol stream derived from bio-diesel production.

2. General considerations

The main objective of this work is the synthesis and economic evaluation of a process for the transformation of crude glycerol, generated as a byproduct in the bio-diesel industry, into liquid hydrocarbons that could be used as transportation fuel. The process synthesis study was conducted using a detailed simulation-evaluation model created in ASPEN PLUS ® - ASPEN ICARUS PROCESS EVALUATOR 2006. This approach allows the use of accurate thermodynamic property calculations, detailed process unit modeling as well as detailed sizing-costing procedures.

2.1. Raw material characterization and plant capacity

The quality of crude glycerol appears to be highly dependent on the facility in which it is produced. A typical crude glycerol mixture coming from the bio-diesel industry includes: 84.5wt% glycerol, 0.3t% methanol, 12.2wt% moisture and 3.0wt% NaCl. The design presented here processes 16,000 ton/y of such mixture.

2.2. Glycerol reforming reactor

The glycerol reformer is an isothermal fixed bed multi-tube reactor. Simonetti et al. (2007) studied the reforming action of a Pt-Re/C catalyst using aqueous solution of glycerol with concentrations between 50wt% and 80wt%, at a temperature of 548 K and pressures between 5 and 17 bar, obtaining good catalyst activities. The associated stoichiometry and kinetic expressions are presented in Table 1. This non-standard ASPEN kinetic model was included in the ASPEN PLUS simulation model as a FORTRAN subroutine.

2.3. Fischer-Tropsch reactor

The FT unit is also an isothermal fixed bed multi-tube reactor. For this study we used the kinetic model for a cobalt-based catalyst as reported by Hamelinck (2003). This kinetic model combines an expression for the total rate of consumption of CO and a product distribution function in the form of the Anderson-Schulz-Flory formula. As before, a FORTRAN subroutine was used to include this non-standard ASPEN PLUS kinetics in the model.

Table 1: Glycerol reforming kinetic model.

$$C_3H_8O_3 \rightarrow 3 \cdot CO + 4 \cdot H_2 \qquad\qquad\qquad\qquad (rxn1)$$

$$H_2O + CO \rightarrow CO_2 + H_2 \qquad\qquad\qquad\qquad (rxn2)$$

$$r_1 = k_1 \cdot \alpha \cdot P_{C_3H_8O_3} \cdot \theta_*^2 \qquad\qquad , \quad r_2 = k_2 \cdot P_{H_2O} \cdot \theta_* - k_{-2} \cdot \theta_{OH} \cdot \sqrt{P_{H_2}}$$

$$\theta_* = \frac{-\left(K_{CO} \cdot P_{CO} + \sqrt{K_{H_2} \cdot P_{H_2}}\right) + \sqrt{\left(K_{CO} \cdot P_{CO} + \sqrt{K_{H_2} \cdot P_{H_2}}\right)^2 + 8 \cdot \alpha \cdot P_{C_3H_8O_3}}}{4 \cdot \alpha \cdot P_{C_3H_8O_3}}$$

$$\theta_{OH} = \frac{k_2 \cdot P_{H_2O} \cdot \theta_* + k_{-3} \cdot P_{CO_2} \cdot \sqrt{P_{H_2}} \cdot \theta_*^2}{k_3 \cdot K_{CO} \cdot P_{CO} \cdot \theta_* + k_{-2} \cdot \sqrt{P_{H_2}}} \qquad k_{-3} = k_2 \cdot k_3 \cdot K_{CO} / \left(k_{-2} \cdot K_{WGS}\right)$$

$$K_{WGS} = 10^{(2073/T[K]-2.029)}$$

$k_i = A_i \, e^{-Ea_i/(RT)}$	A_i		Ea_i	
$k_1 = A_1 \cdot e^{-Ea1/(RT)}$	60·E10	[min⁻¹]	85.6406·E3	[kJ/kmol]
$k_2 = A_2 \cdot e^{-Ea2/(RT)}$	60·E7	[atm⁻¹.min⁻¹]	78.5441·E3	[kJ/kmol]
$k_{-2} = A_{-2} \cdot e^{-Ea-2/(RT)}$	60·E13	[atm⁻¹ᐟ².min⁻¹]	96.3409·E3	[kJ/kmol]
$k_3 = A_3 \cdot e^{-Ea3/(RT)}$	60·E13	[min⁻¹]	55.8972·E3	[kJ/kmol]
$\alpha = A_\alpha \cdot e^{-Ea\alpha/(RT)}$	1·E-7	[atm⁻¹]	-83.7304·E3	[kJ/kmol]
$k_i = A_i \, e^{\Delta H_i/(RT)}$	A_i		ΔH_i	
$K_{CO} = A_{CO} \cdot e^{\Delta Hco/(RT)}$	1E-7	[atm⁻¹]	96.4060E3	[kJ/kmol]
$K_{H2} = A_{H2} \cdot e^{\Delta Hh2/(RT)}$	1E-6	[atm⁻¹]	68.8776E3	[kJ/kmol]

3. Process diagram

The proposed process diagram is shown in **Figure 1**. In this process, crude glycerol is fed to a low pressure evaporation unit to remove inorganic salts that could affect downstream operations (e.g., catalytic reactors). The generated vapor stream is mixed with water to adjust its composition and it is finally condensed. This processed glycerol stream is then compressed and heated before being fed to the glycerol reformer where it is transformed into syngas. The syngas stream is later mixed with a recycle stream of light hydrocarbons and fed to the Fischer-Tropsch reactor to produce a heavier hydrocarbon mixture. Operating conditions were chosen to allow a coupled operation of the FT and GR reactors, eliminating intermediate compression units. The resulting hydrocarbon stream is finally separated into three different cuts: light components, a C_4+ fraction that can be incorporated to transportation fuels, and finally, an aqueous waste stream.

The novelty of the presented process comes from a rhenium-based catalyst (Simonetti et al. 2007) that allows the glycerol reformer to be coupled to the Fischer-Tropsch reactor. This glycerol reformer produces a highly concentrated syngas stream, ready to be used by the FT unit. Operational and capital costs of the reformer are significantly lower than those of biomass gasifiers used in traditional green Fischer-Tropsch processes, leading to a competitive process alternative.

From the thermodynamic point of view, high glycerol conversion and a proper H_2/CO ratio in the syngas produced by the GR reaction are favored by low pressures and high temperatures. On the other hand, high C_{5+} selectivity in the FT reaction is favored by high pressures and low temperatures. These trends do not allow for full heat integration between the two reactors, because thermal coupling of the exothermic FT reactor with the endothermic GR reactor would require the temperature of in the GR reactor to be lower than the temperature in the FT reactor.
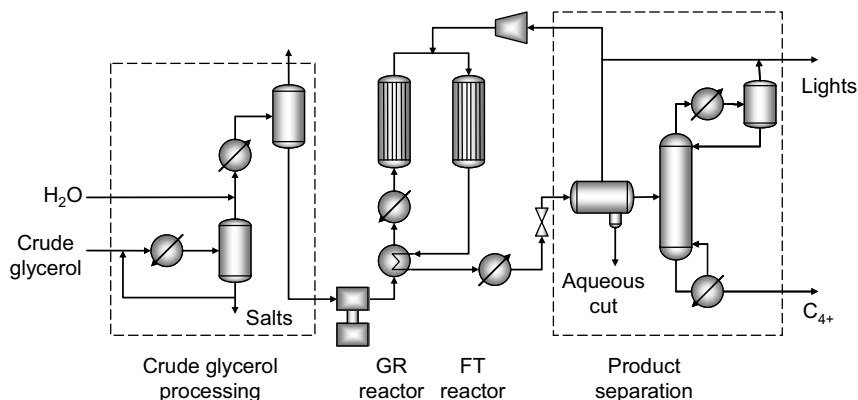
Figure 1: Preliminary glycerol-to-fuel process flow diagram.

However, we found that operating the two reactors at similar pressures while allowing the GR reactor temperature (T_{GRR}) to be higher than the FT reactor temperature (T_{FTR}) is the best option. In this way, an intermediate compression step between the reactors is not required, and the gains from a higher C_{5+} selectivity (resulted from allowing $T_{GRR}>T_{FTR}$) far outweigh the potential savings coming from a reactor heat integration.

For the preliminary process design presented in this study, both GR and FT reactors are operated at approximately 17 bar, while the operating temperatures are 573 and 548 K, respectively. In the FT reactor, these conditions lead to high C_{5+} selectivity. In the GR reactor, these conditions favor the water gas shift reaction, an effect that can be controlled by limiting the water content in the glycerol stream fed to the GR reactor, leading to the proper H_2/CO syngas ratio.

4. Economic evaluation

Detailed economic evaluation of the proposed alternative was conducted using ICARUS PROCESS EVALUATOR 2006 based on the aforementioned detailed process simulation model. Raw material costs are assumed to come solely from crude glycerol, while revenue is considered to come from the sale of light and intermediate liquid hydrocarbon fractions, assuming their values are linked, respectively, to those of natural gas and gasoline. The standard material prices considered here are as follows:

- *Crude glycerol price*: Some reports indicate the high availability of crude glycerol, coming from biodiesel production and other processes, has made its price plunge to levels near 0.02 USD/lb (Yoder et al., 2007).
- *Intermediate hydrocarbon mixture price*: The value of this mixture was considered equivalent to 80% of the current retail prices of gasoline (i.e. 1.80 USD/gal) (http://www.eia.doe.gov/oil_gas/petroleum/data_publications/wrgp/mogas_home_pa ge. html). The current price of oil is 40 USD/bbl.
- *Light hydrocarbon mixture price*: The value of this mixture was assumed to be 30% of the current wellhead price of natural gas (http://tonto.eia.doe.gov/dnav/ng/ng_pri _sum_dcu_nus_m.htm).

The major parameters for this study and the main economic indicators are given in Table 2.

Table 2: Economic evaluation parameters and process evaluation results.

Economic Life [yr]	30	Cap. Cost [USD]	3,899,000
Working Cap./Cap. Expense	5%	Op. Cost [USD/y]	1,436,000
Op. Charged /Op. Labor	15%	Raw Material Cost[USD/y]	644,000
Overhead/Op. Labor	40%	Utility Cost [USD/y]	26,000
Desired RoR [%/yr]	8%	Product Sales [USD/y]	1,957,000
Tax Rate[%/yr]	40%		
Salvage Value/Cap. Cost	20%		
Depreciation	S.L.		
Cap. Escalation [%/yr]	5%		
Material Escalation [%/yr]	1.5%		
Product Escalation [%/yr]	5%		
Utility Escalation [%/yr]	3%		

The Net Present Value (NPV) of the project using the above nominal values is positive ($6.5 \cdot 10^6$ USD). The profitability of the process depends primarily on the price of crude glycerol, which is the sole raw material, and the price of gasoline, which directly determines the price of the major product of the process. To account for uncertainty in these parameters, we carried out sensitivity analysis studies. The goal in these studies was to determine an upper bound on the price of crude glycerol and a lower bound on the price of gasoline that would still allow the processes to be economically feasible (break-even point).

To this end, detailed NPV calculations where performed modifying, one at the time, the two aforementioned prices while keeping all other parameters at their nominal values presented above. The results of these analyses are presented in **Figure 2**. The break-even (maximum) price for glycerol is 4.5¢/lb, well above the current price. Most interestingly, the break-even (minimum) price of gasoline is 1.12$/gal, which is again lower than the current price, indicating that the integrated GR-FT process can be economically attractive even with relatively low oil prices. Thus, we conclude that the proposed process has the economic potential that justifies further development.
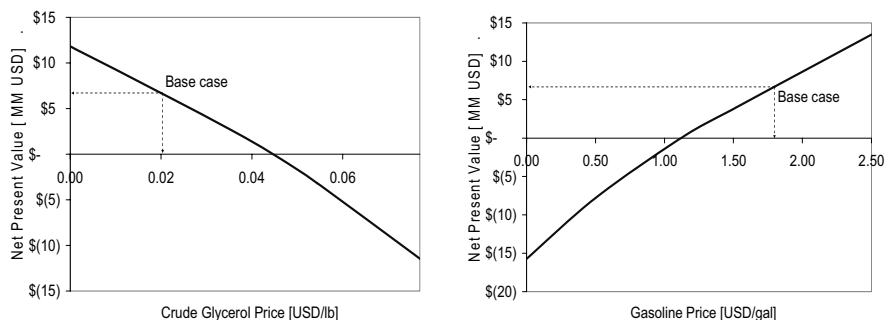
Figure 2: NPV sensitivity analysis for crude glycerol prices and gasoline prices. Nominal values (base case) lead to positive NPV.

5. Conclusions

A novel process for the conversion of crude glycerol to liquid hydrocarbons has been presented and its economic potential evaluated. The proposed design has a number of advantages. First, glycerol reforming (GR) can be carried out at high pressures (up to 17 atm), eliminating the syngas compression step otherwise required to reach FT reaction conditions. Second, glycerol can be reformed at temperatures below 600 K, thus allowing some heat integration with the exothermic FT synthesis. Third, compared to traditional green FT processes (e.g., processes combining biomass gasification and FT), the proposed process features a substantially lower capital cost because it does not require expensive mass gasification and syngas cleaning units. This implies that small- or medium-scale plants can be built closer to biomass production areas. Finally, it allows the use of glycerol coming from other sources such us the fermentation of carbohydrates. Sensitivity analysis on the prices of raw materials and products indicates that the process has the economic potential to justify further development. The proposed process diagram constitutes the result of a preliminary synthesis study. The process can be improved through the development of better catalysts and the use of systematic process synthesis and optimization tools.

References

V.H. Bartholomew, R. J. Farrauto, 2006, Fundamentals of Industrial Catalytic Processes, , Wiley, Hoboken, NJ.

J.J. Bozell, 2006, Feedstocks for the Future using Technology Development as a Guide to Product Identification, ACS Symp. 921, 1-12.

J.N. Chheda, G.W. Huber, J.A. Dumesic, 2007, Liquid-phase Catalytic Processing of Biomass-derived Oxygenated Hydrocarbons to Fuels and Chemicals, Angew. Chem. Int. Ed. 46, 7164-7183.

C. A. Corma, G.W. Huber, L. Sauvanaud, P. O'Connor 2007, Processing Biomass-derived Oxygenates in the Oil Refinery: Catalytic Cracking (FCC) Reaction Pathways and Role of Catalyst, J. Catal, 247, 307-327.

C. N. Hamelinck, A. P. Faaij, H. Uil, H. Boerrigter 2004, Production of FT transsportation fuels from biomass; technical options, porocess analysis and optimisation, and development potential, Energy, 29, 11, 1743-1771.

G.W. Huber, P. O'Connor, C. A. Corma, 2007, Processing biomass in conventional oil refineries: Production of high quality diesel by hydrotreating vegetable oils in heavy vacuum oil mixtures, Appl. Catal. A-Gen, 329, 120-129.

J. Klass, 1998, Biomass for Renewable Energy, Fuels, and Chemicals, Academic Press: San Diego, CA.

A.J. Ragauskas, C.K. Williams, B.H. Davison, et al. 2006, The path forward for biofuels and biomaterials, Science, 311, 484-489.

D. A. Simonetti, J. A. Dumesic 2007, Coupling of Glycerol Processing with Fishcher-Tropsch synthesis for production of liquid fuels, Green Chemistry, 9, 10, 1029-1144.

R.R. Soares, D.A. Simonetti and J.A. Dumesic, 2006, Glycerol as a Source for Fuels and Chemicals by Low-Temperature Catalytic Processing, Angew. Chem. Int. Ed., 45, 3982-3985.

P.L. Spath, D. C. Dayton, 2003, Preliminary Screening--Technical and Economic Assessment of Synthesis Gas to Fuels and Chemicals with Emphasis on the Potential for Biomass-Derived Syngas, United States Department of Energy, National Renewable Energy Laboratory, NREL/TP-510-34929, 2003, 1-142.

J. Yoder, P. Wandschneider (2007). Economics and Policy for Washington State Biofuel Markets: Interim Report. Pg 33.

1725

Dynamic Modeling and Simulation of CO_2 Chemical Absorption Process for Coal-Fired Power Plants

Adekola Lawal[a], Meihong Wang[a], Peter Stephenson[b], Hoi Yeung[a]

[a]*Process Systems Engineering Group, School of Engineering, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom*
[b]*RWE npower, Windmill Park, Swindon, SN5 6PB, UK*

Abstract

Post combustion capture via chemical absorption is viewed as the most mature CO_2 capture technique. The effects of the addition of CO_2 chemical absorption process on power plant performance have been studied using various steady-state models. However, there are several gaps in the understanding of the impact of post combustion capture on the operability of the power plant. These questions could be addressed by studying the dynamic behavior of such plants. In this study, dynamic models of the CO_2 chemical absorption process were developed and validated. Dynamic analyses of the process reveal that absorber performance is sensitive to L/G ratio and that changes in reboiler duty significantly affect the regenerator performance.

Keywords: Post combustion, CO_2 capture, Chemical absorption, dynamic modeling.

1. Introduction

Power generation from fossil fuel-fired power plants is the largest single source of CO_2 emissions [1]. With growing concerns about the environmental impact of such plants, effective CO_2 emission abatement strategies such as Carbon Capture and Storage (CCS) are required for their continued use. One approach to CCS is post combustion capture which involves the separation of CO_2 from the flue gas stream after combustion occurs. Chemical absorption is well suited for separating CO_2 from streams with low concentration of CO_2.

Chemical absorption involves the reaction of CO_2 with a chemical solvent to form a weakly bonded intermediate compound which may be regenerated with the application of heat [2]. Figure 1 describes one of the most popular technologies proposed for post combustion capture. The facility consists of two main units – the absorber and regenerator. Several studies have shown that the energy requirement for solvent regeneration would have adverse effects on power plant efficiency [3]. However, there are several gaps in the understanding of the impact of post combustion capture on the operability of the power plant. For instance, would such power plants be able to effectively operate at varying loads or what modifications are required during start-up [4]? These questions can be addressed by studying the dynamic behaviour of such plants. To achieve this, accurate dynamic models of the power plant and the CO_2 capture facility are required.

2. Developments in modeling chemical absorption of CO_2

Post combustion capture with MEA is a reactive absorption process. Two main phenomena are involved: mass transfer of CO_2 from the bulk vapour to the liquid

solvent and the chemical reaction between CO_2 and the solvent. A number of studies have employed steady state models of the chemical (or reactive) absorption process at different levels of complexity. Kenig et al describes the different levels of complexity of these models [5]. The equilibrium-based approach assumes theoretical stages in which liquid and vapour phases attain equilibrium while the rate-based approach estimates actual mass transfer rates. These models may assume the reactions are at equilibrium or may consider the reaction kinetics. Reaction kinetics could be considered by including an enhancement factor to estimate actual absorption rates (with chemical reactions) from known physical absorption rates. Otherwise, reaction kinetics could be modelled directly [5].
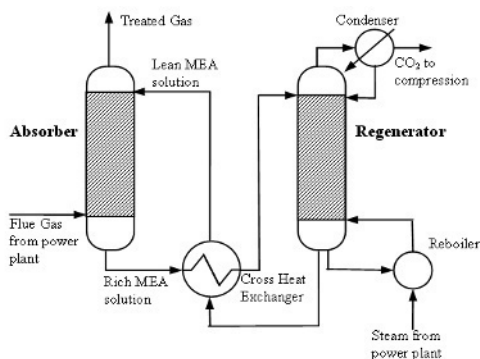
Figure 1 Chemical absorption process for post combustion capture from [4]

The dynamic behavior of the CO_2 absorption process for post combustion capture using MEA has not been extensively studied. Kvamsdal et al considered the dynamic simulation of only the absorber of the process using dynamic models of reduced complexity and also assumed a constant value for the heat of absorption of CO_2 and vaporization of water [4]. Lawal et al considered the dynamic simulation of only the absorber. Mass transfer was based on Maxwell-Stefan formulation and expressions were developed for the heat of absorption of CO_2 [6]. This paper extends the study to the simulation and analysis of both absorber and regenerator columns.

3. Model Development

This section describes the model development of the absorber and regenerator using the rate-based approach for mass transfer.   The physical property method used for both approaches is the Electrolyte Non-random-two-liquid (NRTL) model. MEA electrolyte solution chemistry is used to predict the equilibrium mass fractions in the liquid and vapour phases [7]. Mass transfer is described using the two-film theory (Figure 2) using the Maxwell-Stefan formulation. Heat and mass transfer resistances are modelled in the liquid and vapour films. The rate-based model was developed from the Gas-Liquid Contactor model in Process Systems Enterprise's[1] Advanced Model Library using their process modelling tool, gPROMS. Modifications made include the diffusivity (χ) of CO_2 in the liquid phase which was based on expressions provided by Vaidya et al [8]. The diffusivity (χ) of CO_2 and other components in the vapour phase was estimated using the Fuller's equation [9]. Mass transfer coefficients in the liquid and vapour films were determined by correlations given by Onda et al [10]. Expressions for the heat of absorption and the heat lost to the surrounding were obtained from literature [11,12].

---

[1] Process Systems Enterprise (PSE) Ltd.
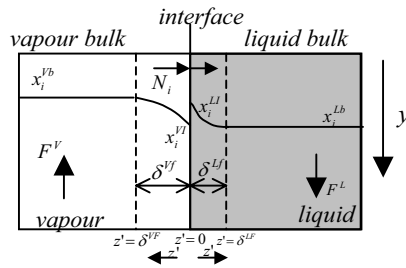
Figure 2 Liquid and vapour bulks, films and interface

3.1. Model Assumptions

The following assumptions were used in developing this dynamic model:

- Plug flow regime and linear pressure drop along the column
- Phase equilibrium at interface between liquid and vapour films
- Negligible solvent degradation

3.2. Model Equations

Mass Balance:
$$\frac{dM_i}{dt} = \frac{-1}{L \cdot A}\frac{\partial F_i^L}{\partial y} + N_i \cdot Sp \cdot MW_i \cdot \omega \tag{1}$$

Energy Balance:
$$\frac{dU}{dt} = \frac{-1}{L \cdot A}\frac{\partial F_H^L}{\partial y} + Sp \cdot \omega \left(H_{liq}^{cond} + H_{liq}^{conv} + H_{abs}\right) + HL \tag{2}$$

Heat of absorption (or desorption):
$$H_{abs} = N_{CO_2} \times h_{abs} \tag{3}$$

Maxwell-Stefan formulation:
$$\frac{1}{\delta}\frac{\partial x_i^M}{\partial z'} = \frac{1}{c_t}\sum_{k=1}^{n}\left(\frac{x_i^M N_k - x_k^M N_i}{\chi_{i,k}}\frac{\mu^R}{\mu}\frac{T}{298.15}\right) \tag{4}$$

The physical property estimation models were set up in Aspen Properties through the CAPE-OPEN Thermo interface.

4. Model Validation

The models developed were validated using data from a pilot plant study [12]. Both absorber and regenerator columns of the pilot plant are packed columns with diameters of 0.427m and total packing heights of 6.1m [12]. Out of the 48 experimental cases carried out in the study, two cases (Cases 32 and 47) were selected for steady state validation purposes. These two cases were selected because of their relatively high and low liquid to gas (L/G) ratios respectively. Simulation results were validated using the temperature profile of both columns measured in the pilot plant [12]. Both columns were simulated separately for validation. In addition the measured CO₂ loading of the amine solvent taken at different positions was compared with values obtained from simulation.

4.1. Case 47

This case involved a relatively low liquid to gas (L/G) ratio in the absorber thus a lower CO₂ capture level. Because of the reported inaccuracy in the flue gas flow measurement [4,12], its value was adjusted to match reported capture levels (Table 1). The temperature profiles in the absorber and regenerator were used to validate the two models as shown in Figure 3(a) and (b). The rate-based model gives fairly good predictions of temperature profiles.
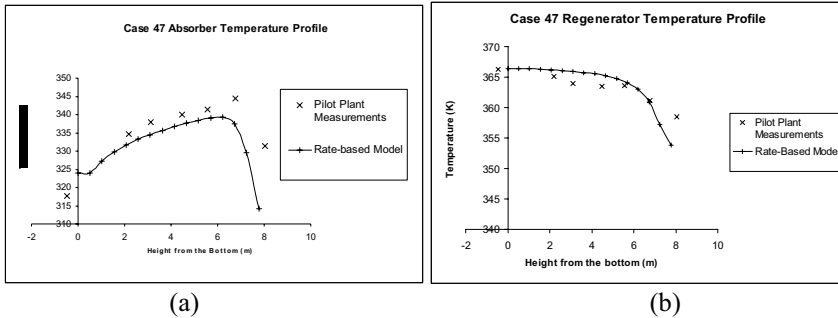
Figure 3 (a) Absorber and (b) Regenerator temperature profile for Case 47

4.2. Case 32

This case involved a relatively high liquid to gas (L/G) ratio thus a high CO_2 capture level. The inlet flue gas flow rate to the absorber was reduced by about 15% to give better predictions of the temperature profile (Figure 4a). However, this change implies higher CO_2 capture levels than what was measured in the pilot plant (Table 1). This discrepancy may be due to the assumption that the reactions between CO_2 and MEA are at equilibrium as calculated by the electrolyte solution chemistry. Kinetically controlled reactions may therefore provide better predictions of the trend.
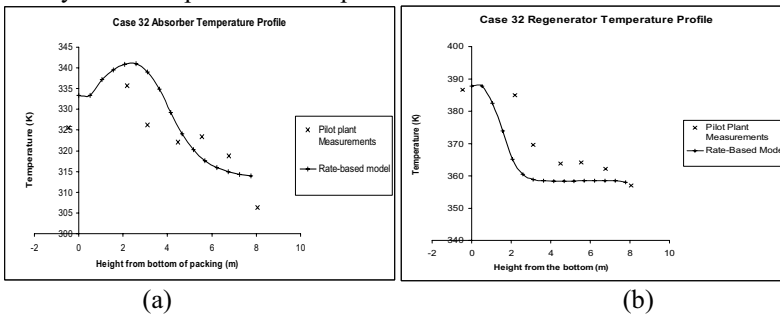
Figure 4(a) Absorber and (b) Regenerator temperature profile for Case 32

Table 1 Comparing CO_2 loading at various locations measured in pilot plant and predicted by rate-based model

	Case 32		Case 47	
	Pilot Plant	Rate Model	Pilot Plant	Rate Model
Absorber Capture Level (%)	95	99.5	69	69.2
Absorber rich MEA loading (mol CO_2/mol MEA)	0.428	0.456	0.539	0.487
Regenerator lean MEA loading (mol CO_2/mol MEA)	0.272	0.260	0.286	0.262

5. Dynamic Analysis

These analyses consider the effect of disturbances on the performance of the columns.

5.1. Reducing Power Plant Load

A 50% reduction in power plant load occurs. Two cases were considered:
- Case-A: Change of flue gas flow rate without changing liquid (solvent) flow rate
- Case-B: Change of flue gas flow rate with corresponding decrease in liquid solvent rate to maintain CO_2 capture level

In Case A, the process was simulated with the base-load conditions (Case 32) for three minutes after which the above changes were implemented in ten minutes. Finally conditions were maintained for eight minutes. From Figure 5a, the 100s curve represents the profile before dropping load. The other curves show a trend of increasing absorption levels with time. Since the flue gas flow rate is ramped down with time while the solvent flow rate is constant, an increase in L/G ratio occurs. In Case B, by reducing the lean solvent feed rate correspondingly (by 50%), roughly the same capture level (Figure 5b) could be maintained through the period of change. This suggests that the absorption process is more sensitive to the L/G (liquid solvent to flue gas) ratio than their actual flow rates. Since the amount of steam required for regeneration corresponds to the amount of lean MEA circulated, the energy requirement of the regenerator could be correspondingly reduced.
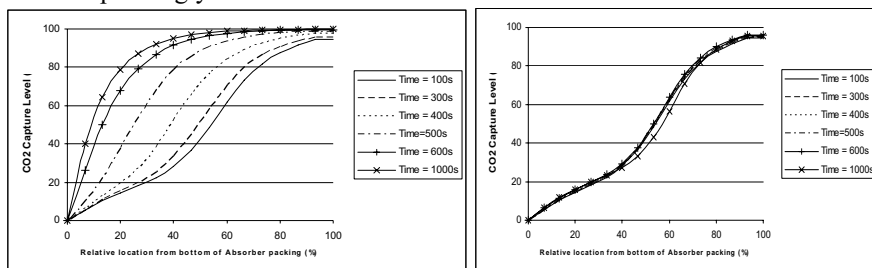
(a) Case A   (b) Case B

Figure 5(a) and (b) Reducing Power Plant Load

5.2. Reducing Regenerator Reboiler Duty
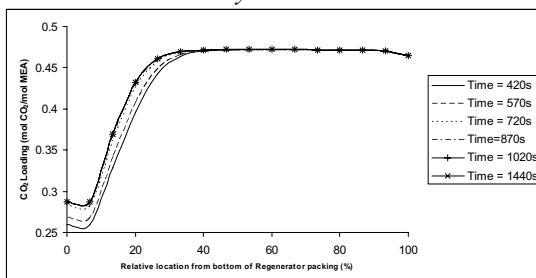
Figure 6 Changes in CO_2 loading profile in regenerator with reducing reboiler duty

The regenerator was simulated with the base-load conditions (Case 32) for seven minutes. A 10% reduction in reboiler duty was implemented over five minutes. Conditions were maintained for 12 minutes. Significant changes in the CO_2 loading profile are observed from the start of the disturbance (time = 420s) onwards especially towards the bottom of the column where CO_2 loading increases (Figure 6). The lean MEA solvent to the absorber would therefore have reduced absorption capacity.

6. Conclusions

This paper presents a study of CO_2 capture using chemical absorption based on the dynamic modelling of the process. Validation results show that the model predicts the absorber and regenerator temperature profiles and CO_2 loadings fairly well. Dynamic analyses show that the absorber performance is more sensitive to the L/G ratio than the actual flow rates of the solvent and flue gas. The performance of the regenerator is significantly affected by the reboiler duty.

Nomenclature

A	Cross sectional area (m^2)	μ	Viscosity (Pa.s)
c_t	Total molar concentration (mol/m^3)	ω	Wetted area ratio
F_i	Component mass flow rate (kg/s)	χ	Diffusivity (m^2/s)
F_H	Enthalpy flow rate (J/s)	*Subscripts*	
H	Heat flux (J/m^2)	*abs*	Absorption
h	Specific Enthalpy (J/kg)	*H*	Enthalpy
HL	Heat loss to surroundings (J/m^2)	*i*	Component number
L	Length of column section (m)	*Liq*	Liquid
L/G	Liquid to gas	*Vap*	Vapour
M	Mass Holdup (kg/m^3)	*Superscripts*	
MW	Molecular weight (kg/mol)	*Cond*	Conduction
N	Molar flux (mol/m^2.s)	*Conv*	Convection
n	Number of components	*I*	Interface
Sp	Specific area (m^2/m^3)	*L*	Liquid
U	Energy Holdup (J/m^3)	*Lb*	Liquid bulk
x	Mass fraction	*Lf*	Liquid film
x_i^M	Molar fraction	*R*	Reference
y	Axial position	*V*	Vapour
z'	Film position	*Vb*	Vapour bulk
Greek Symbols		*Vf*	Vapour film
δ	Film thickness (m)		

References

[1] P. Freund, 2003, Making deep reductions in CO_2 emissions from coal-fired power plant using capture and storage of CO_2, Proc. Inst. Mech. Eng. Part A J. Power Eng.,217(1):1-8.

[2] IPCC, 2005, IPCC special report on carbon dioxide capture and storage.

[3] J. Davison, 2007, Performance and costs of power plants with capture and storage of CO_2, Energy, 32, 7, 1163-1176.

[4] H.M. Kvamsdal, J.P. Jakobsen, K.A. Hoff, 2009, Dynamic modeling and simulation of a CO_2 absorber column for post-combustion CO2 capture. Chemical Engineering and Processing: Process Intensification, 48, 1, 135-144.

[5] E.Y. Kenig, R. Schneider, A. Górak, 2001, Reactive absorption: Optimal process design via optimal modelling, Chemical Engineering Science, 56, 2, 343-350.

[6] A. Lawal, M. Wang, P. Stephenson, H. Yeung, 2008, Dynamic modelling of CO_2 absorption for post combustion capture in coal-fired power plants, Fuel, doi:10.1016/j.fuel.2008.11.009.

[7] Aspentech, 2008, Rate-based model of the CO_2 capture process by MEA using Aspen Plus, Available at: http://support.aspentech.com/. Accessed May, 2008.

[8] P.D. Vaidya, V.V. Mahajani, 2005, Kinetics of the Reaction of CO_2 with Aqueous Formulated Solution Containing Monoethanolamine, N-Methyl-2-pyrrolidone and Diethylene Glycol, Ind. Eng. Chem. Res, 44, 6, 1868-1873.

[9] R.C. Reid, J.M. Prausnitz, T.K. Sherwood, 1977, The properties of gases and liquids, Third ed, New York, McGraw-Hill.

[10] K. Onda, H. Takeuchi, Y. Okumoto, 1968, Mass transfer coefficients between gas and liquid phases in packed columns, Journal of Chemical Engineering of Japan, 1, 1, 56-62.

[11] B.A. Oyenakan, 2007, Modeling of Strippers for CO_2 capture by Aqueous Amines, Doctor of Philosophy Thesis, University of Texas at Austin.

[12] E.R. Dugas, 2006, Pilot plant study of carbon dioxide capture by aqueous monoethanolamine, M.S.E. Thesis, University of Texas at Austin.

Transition to Clean Coal Technologies in India

Anish C. Patil

Delft University of Technology, Faculty of TPM, Delft 2628BX, Netherlands

Abstract

India has the third largest proven reserves of coal in the world and it is obvious that coal will play a major role in its future development. It is well documented that coal is not an environmentally friendly fuel, unless it is utilized in cooperation with clean coal technologies. This paper uses the Transition Management approach to devise a framework for the transition towards clean coal technologies in India. The concept of Transition Management is firmly rooted in the traditions of system thinking that highlights the co-evolution of the socio-technical systems. In order to make technological change sustainable, technical changes alone are not sufficient – changes in the social dimension – such as user practices, governance and institutions are necessary. The role of the government during the transition process is interactive policy formulation – lowering of legal, institutional and policy obstacles and facilitating knowledge transfer and collaboration between the actors.

Keywords: Sustainability, Clean Coal technologies, Transition Management.

1. Introduction

Energy plays a crucial role in the sustainable development of the society – clean, affordable, and uninterrupted supply of energy is the engine for future growth. It is a key ingredient in all sectors of modern economies as they are dependent on an uninterrupted supply of energy. Currently, fossil fuels cater to about 80% of the World's primary energy demand (IEA 2007). The constant dependence on fossil fuels to energize our lives presents a new dilemma – how to satisfy this constantly growing demand for energy given the fact that the fossil fuel resources are finite? The challenge is not just satisfying the increasing energy demand, but at the same time doing it in an environmentally friendly way. When coal, gas and oil are burnt, they release harmful greenhouse gases (GHG) that trap heat in the atmosphere and cause global warming. Because of the long residence times of GHGs in the atmosphere, measured in centuries, climate change would still constitute a risk, as a result of past emissions, even if human-induced emissions suddenly ceased (Kaygusuz 2009). Hence, at this point we cannot stop global warming but only can strive at minimizing its effect for the future generations.

2. Energy demand – India

A rapidly developing country such as India needs greater access to power if they are to achieve the high growth rates required to raise living standards and reduce poverty. On these lines to promote social and economic development, India has invested heavily into building new power plants. As India have large coal reserves, naturally Indian investment in the power sector come along the lines of coal power plants. Coal-based electricity production has driven much of the growth in the Indian electricity sector over the past few decades (COAL 2008). As the future demand for power increases in India,

so will the coal consumption. Although, coal is one of the most abundant energy sources available, it is also the most polluting energy source with very high carbon intensity (Marland and Boden 2001).

The current Indian policies are geared towards increasing the generation capacity that results into the installation of the least risky and cheapest technology. Advanced technologies are necessary for increasing efficiency, reducing environmental pollution and lowering carbon emissions from the Indian power sector (Chikkatur 2008; COAL 2008). India's coal based power generation is primarily based on the subcritical pulverized coal technology – mainly because of the need to rapidly install generation capacity in the country using domestic resources while building up indigenous manufacturing. Indian coal based power plants have low efficiency compared to their counterparts in the developed world. The poor efficiency of India's power plants is usually blamed on a variety of technical and institutional factors such as poor quality of coal, bad grid conditions, degradation due to age, lack of proper operation and maintenance at power plants, ineffective regulations, and lack of incentives for efficiency improvements (Chikkatur 2008). Given the right framework and incentives – Indian power sector too can employ and enjoy clean coal technologies.

The problem at hand is – how can the Indian power sector transition towards clean coal technologies?

Indian power sector is a large socio-technical system that includes, besides the technical subsystem (including the energy carriers, conversion and storage facilities) as well as the people, organizations, and authorities that plan, build, operate, use, and regulate the system (Kaijser 1994; Weijnen and Bosgra 1998). The challenge of transition within a socio-technical system entails joint optimization, at the least coherence, of the social and technical elements of the system (Neij and Åstrand 2006; Ajah, Patil et al. 2007).

Clean coal technologies have the potential to increase the power plant efficiency and are environment friendly, but are not employed on a large scale in India due to the high costs associated with them. Hence, there needs to be proper framework in place that determines which technology is best suited? Who pays for the technology? how long it will take to pay for the technology? etc. For a large country like India how do the policy makers come up with a framework for the transition towards clean coal? This paper suggests that a large developing country like India will need much more than just tweaking around the existing market structure. Just tweaking here and there will not give the long term transition framework that is required for sustainable development. This research proposes transition management approach to draw up a framework for transition towards clean coal for India. 'Transition management' approach can be characterized by the following items:

1. Long-term thinking (at least 25 years) as a framework for shaping the short-term policy thinking, thus trying to bring about system innovation alongside system improvement (Rotmans, Kemp et al. 2001)
2. Multi-level and multi-actor perspective (Geels 2002)
3. A focus on learning and a special learning philosophy (learning-by-doing and doing by-learning) achieved through Bounded Socio Technical Experimentation at the Niche level (Brown, Vergragt et al. 2003)

2.1. Problem background and research question

It is obvious that large developing countries such as India will require huge energy resources to fuel their development further, but at the same time it is paramount to balance this development without compromising the environment. Coal will play a

major part in India's development, but with the threat of climate change looming around it is wise that India invest in clean coal technology otherwise the negative effects on the environment will wipe out the positive economic development. Research question for this paper can be formulated as:

How can the Transition Management approach help to devise a framework for transition towards Clean Coal in India?

3. Transition Management Approach

Transition management aims at understanding socio-technical transitions and at finding methods that facilitate change directed towards achieving societal goals (Geels 2001). The concept "Niches" is of importance for this research, since it defines the level where transitions originate. In order to make changes in the socio-technical system sustainable, mere technological improvements are not sufficient (Rotmans, van Asselt et al. 2000) – changes in the social dimension – such as user practices, rules and regulations are inevitable (Hekkert, Suurs et al. 2007).

The `multi-level' perspective upon which transition management revolves around is shown in figure 1 below. Transition within a socio-technical system distinguishes between the macro-level of the Socio-technical landscape, the meso-level regime, and the micro-level niches.
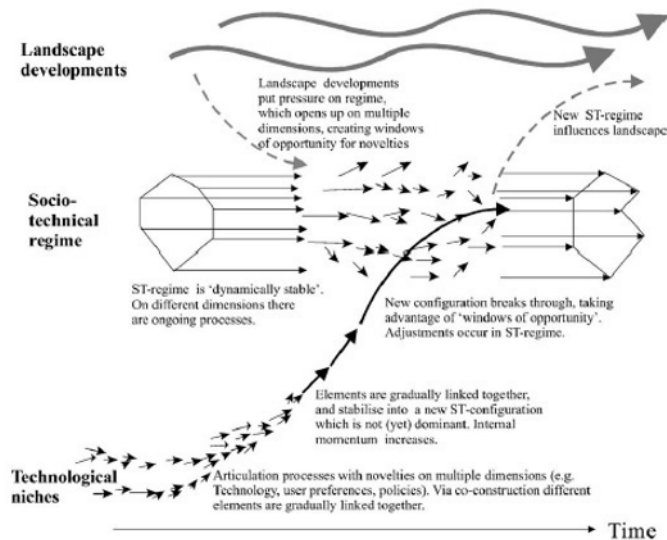
Figure 1: Multi-Level Perspective in Transition Management approach (source: (Geels 2002))

The multi-level approach is centered around the *socio-technical regime* – the socio-technical regime, e.g. energy system, accounts for stability of the socio-technical system through the coordinated and aligned activities of all the actors that are part of the system. The key feature of the 'multi-level perspective' is that system innovations occur through the interplay between the dynamics at these three levels: niche; regime; and landscape (Rotmans, Kemp et al. 2001; Geels 2002). Geels (Geels 2001; Geels 2002) shows that this multi-level model is a nested hierarchy, which can account for the stability in a regime. Geels argues that novelties emerging in technological niches may,

under certain conditions, break through to the regime and come to complement or substitute it – for example, the transition from the usage of horses to cars.

At the highest level of the multi-level model is the socio-technical landscape, which forms an exogenous macro level environment that influences developments in niches and regimes (Geels 2002). The socio-technical landscape tends to change only very slowly (for example, demographic changes, macro-economics, and cultural change). Socio-technical landscape developments refer mainly to national and international developments.

At the lowest level is the Niche level. This is the level that is most dynamic and where novelties emerge. These novelties are initially unstable socio-technical configurations with low performance. Hence, niches act as 'incubation rooms' protecting novelties against mainstream market selection (Kemp 1994; Geels 2002). Studies of past transitions (Kemp 1994; Geels 2002) suggest that a new technology will typically first commercialize in niche markets – where the advantages for this particular technology are the strongest (Utterback 1974; Christensen 1997). Niches help to create virtuous cycles that allow a new technologies to escape lock in, by helping the technology to overcome initial barriers of high costs; the non-availability of complementary technologies; institutional rigidities; and the nonalignment of a new technology to the external environment during the infancy period (Mulder, Reschke et al. 1999).

Niches act as 'incubation rooms' protecting novelties against mainstream market selection (Kemp 1994; Geels 2002). Niches are Bounded Socio-Technical Experiments (BSTE) that exists through large demonstration projects (Brown, Vergragt et al. 2003; Agnolucci and McDowall 2007) – BSTE implies experimentation with a new technology or service, on scale bounded by space and/or time. The time dimension is around five years, but may be longer or shorter, while the space dimension is defined either geographically (a community) or by a number of users (small) (Brown, Vergragt et al. 2003). Studies of past transitions (Kemp 1994; Geels 2002) suggest that a new technology will typically first commercialize in niche markets – where the advantages for this particular technology are the strongest (Utterback 1974; Christensen 1997). Niches help to create virtuous cycles that allow a new technologies to escape lock in, by helping the technology to overcome initial barriers of high costs; the non-availability of complementary technologies; institutional rigidities; and the nonalignment of a new technology to the external environment during the infancy period (Mulder, Reschke et al. 1999). Niches are important for the societal learning and experimentation of a new technology and corresponding social arrangements. The understanding of the underlying mechanisms during the introduction of new technologies in these Niches is of paramount importance for the societal learning.

4. Discussion

Transition towards clean coal technologies in India will encounter many uncertainties. The first step in understanding and managing the transition is to understand the uncertainties in the coal and power sector. To understand these uncertainties extensive literature review was employed (Van der Veen and Wilson 1997; Philibert and Podkanski 2005; Chikkatur 2008; COAL 2008). Figure 2a below classifies the uncertainties at the niche, regime and landscape level on the Y axis, and the level of its impact is shown on the X axis. Such framework is beneficial in understanding the uncertainties at different levels and also the impact they will have on the coal sector. At the macro level oil prices and global energy demand will determine the investment in coal technologies. For example, if oil prices are low then the attention will shift away

from the usage of coal, thus adversely affecting the investments and diffusion of clean coal technologies. At the meso level, environmental standards will play a major role in clean coal technology transition as standards are known to promote technology development (Patil, A.N et al. 2007). At the niche level uncertainties related to the qualities of the technology (such as the price, effectiveness, and availability) are important.

Governance at different levels (macro, meso and niche) for a successful transition towards clean coal is shown in figure 2b below. At the macro level India can participate in the international emissions trading, this will give incentives for the coal sector to reduce emissions. At the meso level, incentives for managers can be provided to improve the productivity of their enterprises and reduce emissions and is thus a precondition for sustainable environmental improvements. The governance should encourage competition and create incentives to improve the efficiency of each sector in the coal energy chain, as well as the efficiency of the chain as a whole (Van der Veen and Wilson 1997). At the niche level there should be policies that will allow easy technology transfer and allow the coal sector to experiment with new clean coal technologies.
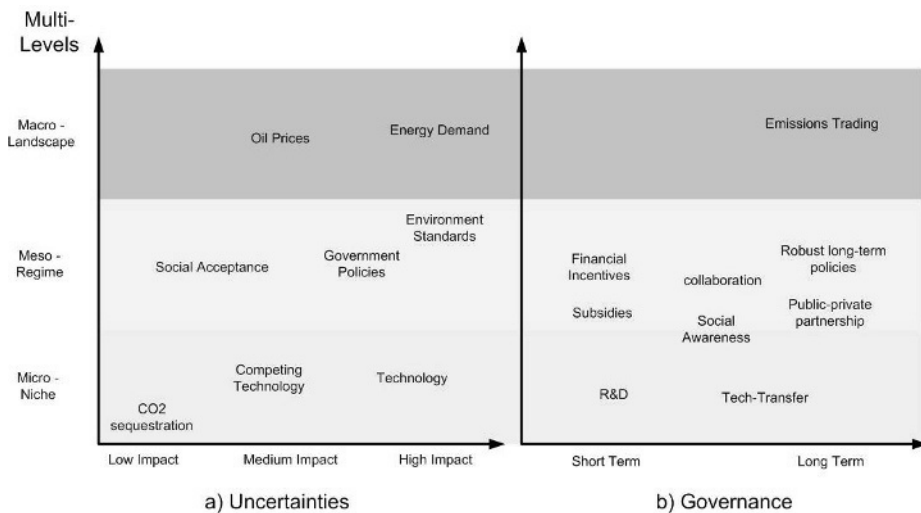
Figure 2: Uncertainties & Governance during the transition towards clean coal

5. Conclusion

The paper concludes that with the coordinated efforts at the national level and good practices at local level the transition towards clean coal technologies can be achieved. Clean coal technologies at the beginning will need protection via subsidies and financial incentives so that they are sustainable in the long run. This paper suggests that the technological improvements alone will not guarantee transition towards clean coal, but such efforts should be complemented by robust and stable policies at the national level. For example: Indian government should sort out the financial incentives for CO_2 sequestration or reduction – this will encourage investments in the environment friendly technologies. To devise the incentives, Indian government could employ the concept of Niches. Niches could entail experimentation at local or project level in order to understand how clean coal technologies can overcome initial barriers of high cost, such

understanding is paramount in formulating a framework for the transition towards clean coal technologies in India.

References

Agnolucci, P. and W. McDowall (2007). "Technological change in niches: Auxiliary Power Units and the hydrogen economy." Technological Forecasting and Social Change 74(8): 1394-1410.

Brown, H. S., P. Vergragt, et al. (2003). "Learning for Sustainability Transition through Bounded Socio-technical Experiments in Personal Mobility." Technology Analysis & Strategic Management 15(3): 291 - 315.

Chikkatur, A. (2008). A Resource and Technology Assessment of Coal Utilization in India, Pew Center on Global Climate Change.

Christensen, C. (1997). The Innovator's Dilemma. Cambridge, Harvard Business School Press.

COAL (2008). Annual Report Ministry of Coal - Government of India.

Geels, F. (2001). Technological Transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study. Nelson and Winter Conference, Aalborg, Denmark.

Geels, F. (2002). Understanding the Dynamics of Technological Transitions. Twente, Twente University Press.

Hekkert, M. P., R. A. A. Suurs, et al. (2007). "Functions of innovation systems: A new approach for analysing technological change." Technological Forecasting and Social Change 74(4): 413-432.

IEA (2007). Renewables in Global Energy Supply - An IEA Factsheet. Paris, International Energy Agency.

Kaijser, A. (1994). I fädrens spår: Den svenska infrastrukturens historiska utveckling och framtida utmaningar. Stockholm, Sweden, Carlsson.

Kaygusuz, K. (2009). "Energy and environmental issues relating to greenhouse gas emissions for sustainable development in Turkey." Renewable and Sustainable Energy Reviews 13(1): 253-270.

Kemp, R. (1994). "Technology and the transition to environmental sustainability : The problem of technological regime shifts." Futures 26(10): 1023-1046.

Marland, G. and T. Boden (2001). The Increasing Concentration of Atmospheric CO2: How Much, When, and Why? Erice International Seminars on Planetary Emergencies, 26th Session, Sicily, Italy.

Mulder, P., C. H. Reschke, et al. (1999). Evolutionary Theorising on Technological Change and Sustainable Development. European Meeting on Applied Evolutionary Economics. Grenoble, France, Institute for Energy Politics and Economics.

Patil, A. C., A. A.N, et al. (2007). Trends in Emission Standards for Public Transport Buses: How Low should they go. IAEE Asian Conference: Asian Energy Security and Economic Development in an Era of High Oil Prices, Taiwan.

Philibert, C. and J. Podkanski (2005). Clean Coal Technologies. International Energy Technology - Collaboration and Climate Change Mitigation, International Energy Agency.

Rotmans, J., R. Kemp, et al. (2001). "More evolution then revolution : transition management in public policy." Foresight - the journal of foresight studies, strategic thinking and policy 3(1).

Rotmans, J., M. van Asselt, et al. (2000). "Visions for a sustainable Europe." Futures 32(9-10): 809-831.

Utterback, J. M. (1974). "Innovation in Industry and the Diffusion of Technology." Science 183(4125): 620-626.

Van der Veen, P. and C. Wilson (1997). "A new initiative to promote clean coal." Finance and Development 34(4): 36-39.

Weijnen, M. and O. Bosgra (1998). An Engineering Perspective on the Design and Control of Infrastructure - Explorations into a Generic Approach to Infrastructure Scenario Analysis. The Infrastructure playing field in 2030. M. P. C. Weijnen and E. F. t. Heuvelhof. Noordwijk, Delft University Press.

1737

Storage Logistics of Fruits and Vegetables in Distribution Centers

Daniela F. Borghi[a], Reginaldo Guirardello[a], Lúcio Cardozo Filho[b]

[a] School of Chemical Engineering, State University of Campinas – UNICAMP – P.O. Box 6066, 13083-970, Campinas – SP, Brazil
[b] Department of Chemical Engineering, University of Maringá – UEM – Colombo Avenue, 5790 – 87020-900, Maringá – PR, Brazil

Abstract

This study has as main objective to minimize the losses of fruits and vegetables that happen during its storage in distribution centers, using mathematical models to optimize the products distribution in deposit and to minimize costs referring to the storage of these products. Data of keeping quality of some horticultural products were taken from the literature. The mathematical model for optimization used a mixed integer linear programming (MILP) formulation. It was implemented in the software GAMS® and the solver used was CPLEX®. Simulations were conducted considering various temperatures of storage and the same cost of U$ 1.00 for every analyzed products (celery, lettuce, leek, plum, beetroot, onion, carrot, chicory, cauliflower, spinach, papaya, strawberry, turnip, cucumber, peach, bell pepper, radish, parsley, tomato and grape), aiming to minimize the costs relating to its storage. Starting from values of $KQ_{ij(real)}$ (keeping quality of product i in zone j in real storage temperature) and $KQ_{i(ideal)}$ (keeping quality of product i in its ideal storage temperature) and considering two zones (refrigerating chamber and external deposit) it was possible to proceed an analysis in relation to keeping quality of each product in function of the temperature. The results showed that the more distant the real storage temperature of a product is of its ideal temperature, the greater the costs with the loss of quality of this product. It was verified that the cost relating to the loss of quality of products has great influence on the total storage cost. It shows the importance of optimizing the storage logistics of these products, so that these costs are not still more expressive, what certainly occur without a tool for that. It was also verified that the order of products priority to occupy the chamber depends on the zones temperatures combinations, but fruits as grape and plum have larger priority in majority combinations of temperatures, followed by peach and strawberry. Papaya, bell pepper, tomato and cucumber have great variation in its priority and for smaller chamber temperatures (T_c), the products with smaller priorities were cucumber and tomato and for larger T_c were cauliflower and celery. In that way, the proposed model comes as a tool that can help the distribution centers managers in taking decisions referring to the logistics of storage of fruits and vegetables in order to minimize total costs and losses of the sector.

Keywords: fruits and vegetables, keeping quality, cost of storage, optimization.

1. Introduction

The storage of fruits and vegetables has a great importance to the establishments that commercialize it, because they are very perishable and sometimes become a detriment to the establishment, due to the great quantity of product waived because of loss of

quality. To get an idea of the size of this problem in Brazil, in 2001, were harvested 15 million tons of vegetables, and more than 5 million tons were lost, enough to supply 53 million people and generate a loss more than US$ 1 million (Vilela et al., 2003).

Fruits and vegetables are susceptible to loss of keeping quality due to injuries or loss of moisture. The keeping quality of a product is the time until it becomes unacceptable for consumption, from the sensory, nutritional or safety point of view (Fu and Labuza, 1993).

The storage temperature is one of the most important factors that influence the keeping quality of fruits and vegetables. Low temperatures reduce the respiration rate, the speed of growth of pathogenic microorganisms and loss of moisture, and delay the senescence and changes in color and texture (Borghi, 2008).

It is known that the management of a deposit of fruits and vegetables is more complex than a deposit of processed products as the first ones need rapid commercialization and special conditions of storage to reduce its loss of quality, because it reflects in reduction of its commercial value.

The main objective of this study is to verify the influence of temperature on the storage logistics of fruits and vegetables in distribution centers, using a mathematical model to optimize the distribution of the products and to minimize the costs related to its storage.

2. Mathematical Model

To develop the mathematical model were needed some data describing the physical space available, the volume of product to be stored, the conditions under which products must be kept and the location of these within the deposit. Knowing this information, it was possible to determine which area in the accommodation of the deposit, with specific conditions of storage that a product should take. This area was called "zone" by Broekmeulen (1998). It was considered that in a distribution center deposit there are basically two zones, the external deposit, subject to room temperature, and the refrigeration chamber, maintained at a pre-established temperature.

Three assumptions were considered. The first one is that the storage time and temperature are the only storage conditions that influence the reduction of products keeping quality. The second is that the storage temperature depends on the zone. And the last one is that the stock levels and the models of keeping quality change are known for all products.

Define the decision variable x_{ij} (binary variable) for all $i \in P$ and $j \in Z$:

$$\begin{cases} x_{ij} = 1 \text{ if the product } i \text{ is designated to zone } j \\ x_{ij} = 0 \text{ in any other case} \end{cases}$$

Thus, the problem may be placed in the form of a mixed integer linear programming (MILP) (Weng, 2008). Then, the problem is to find a plan for distribution of fruits and vegetables that minimizes the cost function (C) given by:

$$C = \sum_{i \in P} \sum_{j \in Z} \frac{c_i . s_i}{d_{ij}} x_{ij} + \sum_{i \in P} \sum_{j \in Z} \frac{b_{ij} . c_i . s_i}{d_{ij}} x_{ij} + \sum_{i \in P} \sum_{j \in Z} \frac{q_{ij} . c_i . s_i}{d_{ij}} . x_{ij} + \sum_{j \in Z} a_j . S_j$$

(1)

and fulfilling the following restrictions:
1. Each product can be assigned to only one specific zone.
2. The storage capacity of the zone can not be exceeded.

3. The time that the product i stays in zone j can not exceed the keeping quality of this product.

The cost function (C) depends on several factors, the parameters of the model. The set P includes all fruits and vegetables studied and each receives an index i. The set of zones (Z) encompasses all zones j, each with specific conditions and a physical capacity of storage (S_j).

The parameter a_j is the storage cost of zones. Keep a cold chamber requires an expense of extra energy compared to the deposit. For this reason, the costs to storage of the products in the deposit and the cold chamber are different. The percentage of products loss (b_{ij}) depends on many factors such as transport, handling of the products and inadequate storage, which added, lead to loss of the products. The acquisition cost of the products (c_i) is the price paid per kilo or unit of the product. The parameter d_{ij} is the time that the product is stored, which was considered equal to a day for all products.

Another parameter of the model is q_{ij} (fraction of the original quality that is lost when compared with the storage in ideal conditions for a product). This parameter is obtained through the following relationship:

$$q_{ij} = 1 - \frac{KQ_{ij(real)}}{KQ_{i(ideal)}}$$
(2)

where $KQ_{ij(real)}$ and $KQ_{i(ideal)}$ is the keeping quality of product i in the real storage temperature and the keeping quality of product i in its ideal temperature, respectively.

Tijskens and Polderdijk (1996) developed a useful formulation for calculating the keeping quality of fruits and vegetables. They showed that the keeping quality is proportional to the inverse of reaction rate k, independent of the kinetic mechanism of the loss of product quality. Thus, it is possible to describe the behavior of keeping quality as a function of temperature. To Labuza (1984), the reactions of loss of quality of food can be described by the Arrhenius Law, where specific reaction rates depend on temperature. The reaction rate k can be approximated by this Law:

$$k = k_{ref} e^{\frac{E_A}{R}\left(\frac{1}{T_{ref}} - \frac{1}{T_{abs}}\right)}$$
(3)

In many fruits and vegetables, the quality attribute which limits consumers acceptance, changes at a different temperature. This can be seen in cold sensitive products. In tomatoes, for example, kept at constant temperatures below 8 °C, the limiting factor is usually color, but when the temperature is above 13 °C, the limiting factor is firmness.

The keeping quality proposed by Tijskens and Polderdijk (1996) under constant conditions, may be represented by:

$$KQ = \frac{KQ_{ref}}{\sum_{n=1}^{N} k_{ref(n)} e^{\frac{E_{A(n)}}{R}\left(\frac{1}{T_{ref}} - \frac{1}{T_{abs}}\right)}}$$
(4)

where N is the process number that contribute to the keeping quality, which normally is not greater than two.

Assuming $k_{ref(1)}$ (reaction rate of the first component in the reference temperature) equal to one, statistical analysis was made possible without the need for information on the

type of quality function. Moreover, the expression for the keeping quality becomes almost independent of the type of kinetic mechanism involved.

The loss of product quality is usually measured by analysis of one or more parameters which may be physical, chemical, microbiological or sensorial (Taoukis and Labuza, 1989). It is possible to determine the keeping quality of a product at a reference temperature (KQ_{ref}) using methods for assessing quality, which may be measurement of color and firmness, for example (Mitcham et al., 1996).

This method is generic and applicable to products of moderate and tropical areas. This model can be used, therefore, to predict the keeping quality of fruits and vegetables stored in a supermarket. The proposed equation was implemented in the software GAMS® and optimizations were made leading into account different temperatures of the cold chamber and the deposit and variability in quantities of products stored in the distribution center, using the CPLEX® solver.

3. Results and Discussion

Optimizations were performed using Equation (1) and the restrictions, taking into account different temperatures for zones and different quantities of products purchased by the distribution center. Initially, the values of $KQ_{ij(real)}$ and $KQ_{i(ideal)}$ were obtained using Equation (4) and data estimated by Tijskens and Polderdijk (1996). It was found that for products with a number of process equal to one ($N = 1$) the keeping quality increases as temperature decreases. However, for products with number of process equal to two ($N = 2$), both in low as in high temperature, the keeping quality is small and increases as it approaches the ideal temperature of storage.

From these data was obtained the change of the quality factor, q_{ij}, with the temperature by Equation (2). The calculated results are shown in Figure 1.

The analysis of results shows that when the real storage temperature of each product is equal to its ideal temperature, $q_{ij} = 0$, there is no cost to the loss of quality of this product.

The maximum value of q_{ij} is 1 and the higher its value, more distant of ideality a product is stored and the greater the costs for its loss of quality.

The priority order of the studied products to stay in the cold chamber was also evaluated. For that, it was considered that the amount of each product was equal a box and the capacity of the chamber was increased one by one box to see what would the first product in the chamber. This procedure was performed for many combinations of chamber temperature (T_c) and stock temperature (T_e), with T_c of 5 °C to 15 °C and T_e of 10 °C to 33 °C, both ranging from one by one degree. The price paid for each product used in these optimizations was U$ 1.00 for all products.

Analyzing the results, it was observed that the grape, the first in order of priority in lower T_c, loses its place for the papaya, as T_c increases. The plum, peach and strawberry have its priority reduced because papaya and tomatoes pass to enter before in the chamber in higher T_c (T_c is approaching the T_{ideal} of these fruits).

The radish loses priority for the onion, bell pepper and cucumber, which pass to enter before in the chamber in higher T_c. The spinach lost its place to turnip and parsley to spinach, turnip and beetroot in higher T_c.

The results described above do not take into account the third restriction. According to this restriction, the keeping quality of the product i in zone j must be greater than a minimum period of residence of this product in this zone. So, for some values of T_e some products should necessarily be stored in cold chamber, as quickly ruined in stock, taking its keeping quality lower than d_{ij}. Optimizations were carried out taking into

account the third restriction. The results show that for a T_e up to 17 °C, independent of T_c, there is no product that is unable to remain in stock. In these situations, the results are very similar to those obtained in the optimizations that do not consider this restriction. For other cases the results are presented in Table 1.
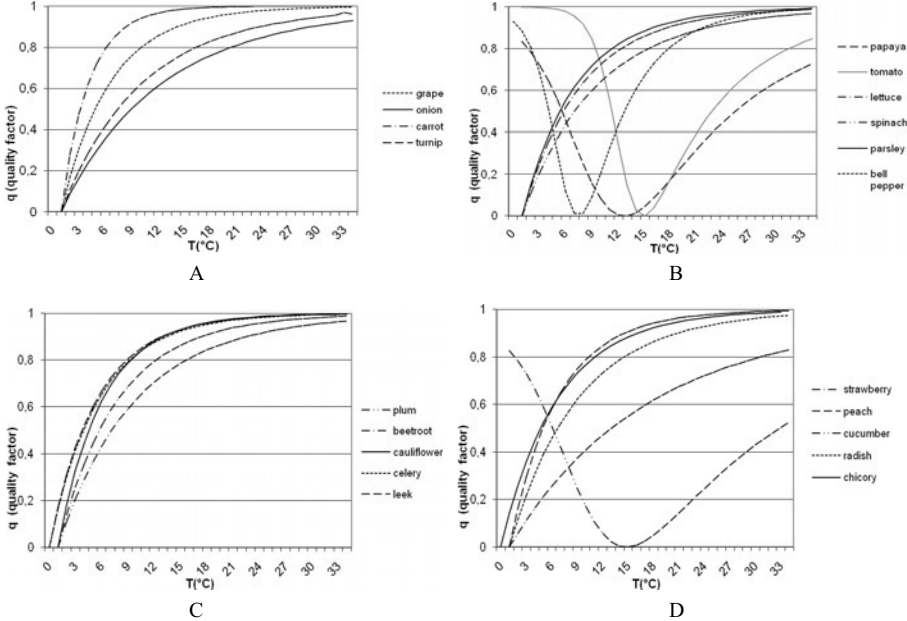
Figure 1. Change in quality factor of products with temperature. For all the Figures (A, B, C and D), when the product is stored exactly at its ideal temperature, q_{ij} is equal to zero, and it value increases as more distant is the storage temperature of ideal temperature. Figures A and C: the storage ideal temperature of all the products is 0 °C. Figure B: the storage ideal temperature of papaya, tomato and bell pepper is 12 °C, 14 °C and 7.5 °C, respectively, and for the other products it is equal to 0 °C. Figure D: the storage ideal temperature of cucumber is 14 °C and for the other products it is equal to 0 °C.

Table 1. Products that should be stored in cold chamber due to the third restriction.

T_e	Products that should be stored in cold chamber
18	peach, lettuce, carrot, radish
20	peach, lettuce, carrot, spinach, radish, celery, leek
22	peach, lettuce, carrot, cauliflower, spinach, radish, parsley, celery, leek
24	peach, lettuce, beetroot, carrot, cauliflower, spinach, radish, parsley, celery, chicory, leek, bell pepper
26	strawberry, peach, lettuce, beetroot, carrot, cauliflower, spinach, radish, parsley, celery, chicory, leek, bell pepper
28	strawberry, peach, lettuce, beetroot, carrot, cauliflower, spinach, radish, parsley, celery, chicory, leek, bell pepper
30	plum, strawberry, peach, lettuce, beetroot, carrot, cauliflower, spinach, radish, parsley, celery, chicory, leek, bell pepper
32	plum, strawberry, peach, lettuce, beetroot, carrot, cauliflower, spinach, radish, parsley, celery, chicory, leek, bell pepper
33	plum, strawberry, peach, grape, lettuce, beetroot, carrot, cauliflower, spinach, radish, parsley, celery, chicory, leek, bell pepper

Looking at the table above and the Figure 1, it is possible to observed that the order of entry of products into the cold chamber is affected by the third restriction and by the temperatures T_c and T_e. The importance of Restriction 3 is perceived, as if it were ignored many products remain in stock, even taking its keeping quality in that zone ($KQ_{ij\ (real)}$) less than the minimum keeping quality necessary for this product (d_{ij}).

In the optimizations performed in this work, the parameter d_{ij} was considered equal to one for all products, this means that the minimum time of residence of all the products in any of the zones should be one day. It was noted that, for example, lettuce, cauliflower and celery, which have the keeping quality less than a day in stock even for T_e not too high, are some of the latest products in order of priority to remain in the cold chamber when optimizations were performed not considering this restriction.

4. Conclusions

The proposed model is a tool that helps reduce the total cost of the establishment related to the storage of fruits and vegetables.

The equation proposed by Tijskens and Polderdijk (1996) was a useful tool to predict the keeping quality of the products in different storage temperatures.

It was found that the temperature in the chamber and the temperature of the stock significantly affect the order of priority to the products be in the cold chamber and that is very important to consider the keeping quality of each product in a certain temperature before deciding on the logistics of these products within the deposit.

The method is generic and therefore can be used as a tool to assist in decision making related to the logistics of storage of fruits and vegetables in supermarket.

References

D. F. Borghi, 2008, Logística de armazenamento de frutos e hortaliças em supermercado, 203p. Dissertation (Masters in Chemical Engineering) – School of Chemical Engineering, State University of Campinas, Campinas.

R. A. C. M. Broekmeulen, 1998, Operations Management of Distribution Centers for Vegetables and Fruits. International Transactions in Operational Research. vol. 5, n. 6, 501-508.

B. Fu; T. P. Labuza, 1993, Shelf-life Prediction: Theory and Application. Food Control, Butterworth-Heinemann, vol. 4, n. 3, 125–133.

T. P. Labuza, 1984, Application of Chemical Kinetics to Deterioration of Foods. Journal of Chemical Education, vol. 61, n. 4.

B. Mitcham; M. Cantwell; A. Kader, 1996, Methods for Determining Quality of Fresh Commodities. Perishables Handling Newsletter Issue, n. 85.

P. S. Taoukis; T. P. Labuza, 1989, Applicability of time-temperature indicators as shelf-life monitors of food products. Journal of Food Science, vol. 54, n. 4, 783–788.

L. M. M. Tijskens; J. J. Polderdijk, 1996 A Generic Model for Keeping Quality of Vegetable Produce During Storage and Distribution. Agricultural Systems, vol. 51, n. 4, 431–452.

N. J. Vilela; M. M. Lana; E. F. Nascimento; N. Makishima, 2003, O peso da perda de alimentos para a sociedade: o caso das hortaliças. Hortic. Bras. vol. 21, n. 2, 142–144.

M. Weng, Integer Programming. In: A. R. Ravindran, *Operations Research and Management Science Handbook.* CRC Press, 2008.

Optimal Control Law Development in a Sequential Batch Reactor Through Mixed Integer Particle Swarm Dynamic Optimization

Adrián Ferrari[a], Soledad Gutierrez[a], Evaristo C. Biscaia[b]

[a]*IIQ - Department of Reactors Engineering, Engineering School - University of the Oriental Republic of Uruguay, Julio Herrera y Reissig 565, CP 11300, Montevideo, Uruguay.*
[b]*LMSCP/PEQ/COPPE/UFRJ, Av. Horácio Macedo, 2030 - Centro de Tecnologia - Bloco G - Sala G-115 - CEP 21941-914 - Rio de Janeiro – RJ, Brazil.*

Abstract

Dynamic optimization in Sequential Batch Reactors for wastewater treatment represents an enormous challenge in order to improve the time and energy management in real industries. The non convex behaviour presented by these systems limits the application of deterministic techniques to optimize this kind of equipment. Although any real example has been found in the open literature, stochastic contributions to meet global optimization goals in the SBR technology appear to be a promising strategy. A Particle Swarm Optimization (PSO) algorithm in order to minimize the total aeration demand (energy management) in a real dairy SBR process (operating in an Intermittent Aeration Extended Filled configuration) for carbon and nitrogen removal was here developed. The batch reactor network size, sequencing and stages duration, were assumed as the decision variables for the dynamic MINLP problem. Two kinds of PSO algorithms (relaxed; and mixed-integer) were here tried in order to find the best way to take into account the mixed-integer nature of the system. In addition, PSO results were compared with the ones obtained by a sequential shooting method/NLP Code, considering different initial guesses. Stochastic optimization allowed improving the obtained results from deterministic techniques with a very reasonable computational cost, and mixed-integer PSO resulted in the best structure solving the MINLP problem. Despite of that, and in order to assure the most robust and reliable results for this optimization problem (high swarm convergence and good optimization efficiency), the assessment of both PSO formulations must be considered. The PSO results have given an optimal control law that minimizes the total aeration demand along the SBR operation; this law can be easily implemented (switching times).

Keywords: SBR, Dynamic Optimization, Particle Swarm Optimization.

1. Introduction

Sequential Batch Reactors (SBR) represents one of the most important activated sludge technologies for wastewater (WW) treatment. Carbon, nitrogen and phosphorous removal can be carried out simultaneously in this kind of equipment (Artan and Orhon; 2005). The extreme flexibility presented by these systems [feeding pattern and reaction network establishment (Ferrari et. al[1]; 2008)] encourages the use of optimal control laws even during its operation at industrial scale. This feature is rarely observed for the rest of WW treatment technologies. Figure 1 shows an example of a usual SBR process treating continuous WW for carbon and nitrogen removal.
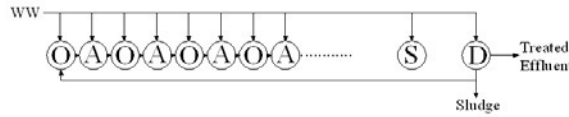
Figure 1: Intermittent Aeration Extended Filled (IAEF) Reactor (Artan and Orhon, 2005). Stages: **O**, Oxic (Aerobic); **A**, Anoxic; **S**, Settling; and **D**, Draw.

Dynamic optimization in Sequential Batch Reactors represents an enormous challenge in order to improve the time and energy management in real WW treatment plants. The non convex behaviour presented by these systems limits the application of deterministic techniques to optimize this kind of equipment. Although any real example have been found in the open literature (Souza et. al; 2008, Coelho et. al; 2001), stochastic contributions to meet global optimization goals in the SBR technology appears to be a promising strategy. Very few papers related with energy optimization in this kind of equipment could be found. Despite the time management represents the usual objective function for the optimization problem in Sequential Batch Reactors (Souza et. al; 2008, Coelho et. al; 2001, Ferrari et. al[2]; 2008), new insights of optimal control laws can be obtained through the analysis of their energy management. The main purpose of this contribution is to develop an optimal control law in order to minimize the total aeration demand (energy management) along the operation of a real SBR process, through the application of Particle Swarm Optimization (PSO) (Kennedy and Eberhart; 1995). In addition, a comparison between PSO results with the ones obtained from a sequential shooting method/NLP Code, in order to verify the non convex behaviour of the system, is presented in this paper.

2. Mathematical Modelling and Applied Methods

2.1. SBR Model and Kinetic Description

The dimensionless transient mass balance applied to each component in the system is presented below (lumped structure):

$$\frac{dc_i(t)}{dt} = Fe_{SBR} \cdot \frac{q(t)}{v(t)} \cdot \left[c_{f,i}(t) - c_i(t) \right] - Da_i \cdot r_i \left[\boldsymbol{c}(t) \right]$$

Where c_i is the component i concentration; $c_{f,i}$, the component i feed concentration; t, the time variable; q, the volumetric wastewater flow rate; v, the reactor volume; r_i, the component i consumption rate; $\boldsymbol{Da_i}$, the Damköhler number for component i; $\boldsymbol{Fe_{SBR}}$, the Feeding number; and $\boldsymbol{c}$, the column vector of the concentration of all components. More details regarding dimensionless SBR models are presented in Ferrari et. al[1] (2008). Aerobic/anoxic carbon removal and autotrophic nitrification represents the main processes involved in this equipment. The biomasses growth kinetics adopted in this work are in accordance to ASM1 Model (Orhon and Artan; 1994) and their complete description is presented below.

Anoxic Carbon Removal (Denitrification)

$$r_{x_H} = (\mu_{hN} - kd_{hN}).X_H$$

$$r_{s_C} = \frac{1}{Y_{hN}}.\mu_{hN}.X_H \qquad \mu_{hN} = \mu_{hNmax}\left(\frac{S_C}{K_{C1}+S_C}\right)\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right) \qquad r_{s_{NO}} = \frac{1}{Y_{\frac{X}{N}}}.\mu_{hN}.X_H + kd_{hN}.X_H.f_{\frac{N}{X}}\left(\frac{S_{NO}}{K_{NO}+S_{NO}}\right)$$

Aerobic Carbon Removal

$$r_{x_H} = (\mu_h - kd_h).X_H \qquad \mu_h = \mu_{hmax}\left(\frac{S_C}{K_{Ch}+S_C}\right)\left(\frac{S_O}{K_O+S_O}\right) \qquad r_{s_C} = \frac{1}{Y_H}.\mu_h.X_H$$

Autotrophic Nitrification

$$r_{x_A} = (\mu_a - kd_a).X_A \qquad \mu_a = \mu_{amax}\left(\frac{S_{NH}}{K_{NHa}+S_{NH}}\right)\left(\frac{S_O}{K_{Oa}+S_O}\right) \qquad r_{s_{NH}} = \frac{1}{Y_{NH}}.\mu_a.X_A \qquad r_{s_{NO}} = -\frac{1}{Y_{NO}}.\mu_a.X_A$$

Table 1: Nomenclature Description for Kinetic Expressions

Symbol Components	Name	Value	Dimension
S_C	Soluble Chemical Oxygen Demand ($COD_{Soluble}$)	-	$gCOD.m^{-3}$
S_{NO} / S_{NH}	Nitrate (NO_3^-) / Ammonium (NH_4^+) Concentration	-	$gN.m^{-3}$
S_O	Dissolved Oxygen (DO)	-	$gO_2.m^{-3}$
X_H / X_A	Heterotrophic / Autotrophic Biomass Concentration	-	$gVSS.m^{-3}$
Kinetic Parameters			
μ_{hmax}	Maximum Specific Aerobic Growth Rate for X_H	1	d^{-1}
μ_h	Specific Aerobic Growth Rate for X_H	-	d^{-1}
K_{Ch}	Aerobic Saturation Constant for S_C	75	$gCOD.m^{-3}$
K_O / K_{Oa}	Saturation Constant for S_O	~ 0	$gO_2.m^{-3}$
kd_h / kd_a	Aerobic Decay Coefficient for X_H / X_A	0.05 / 0.025	d^{-1}
μ_{hNmax}	Maximum Specific Anoxic Growth Rate for X_H	0.75	d^{-1}
μ_{hN}	Specific Anoxic Growth Rate for X_H	-	d^{-1}
K_{Cl}	Anoxic Saturation Constant for S_C	75	$gCOD.m^{-3}$
K_{NO}	Saturation Constant for S_{NO}	15	$gN.m^{-3}$
kd_{hN}	Anoxic Decay Coefficient for X_H	0.05	d^{-1}
μ_{amax}	Maximum Specific Aerobic Growth Rate for X_A	0.5	d^{-1}
μ_a	Specific Aerobic Growth Rate for X_A	-	d^{-1}
K_{NHa}	Aerobic Saturation Constant for S_{NH}	1	$gN.m^{-3}$
Yield Parameters			
Y_H	Aerobic Heterotrophic Biomass Yield Coefficient	0.42	$gVSS.gCOD^{-1}$
Y_{NH}	Aerobic Autotrophic Biomass Yield Coefficient	0.170	$gVSS.(gN-NH_4^+)^{-1}$
Y_{NO}		0.174	$gVSS.(gN-NO_3^-)^{-1}$
Y_{hN}	Anoxic Heterotrophic Biomass Yield Coefficient	0.42	$gVSS.gCOD^{-1}$
$Y_{X/N}$		1.67	$gVSS.(gN-NO_3^-)^{-1}$
$f_{N/X}$	Chemical Nitrate Demand for X_H	0.36	$(gN-NO_3^-).gVSS^{-1}$
Reaction Rates			
r_{XH} / r_{XA}	Growth Rate for X_H / X_A	-	$gVSS.m^{-3}.d^{-1}$
r_{SC}	Removal Rate for S_C	-	$gCOD.m^{-3}.d^{-1}$
r_{SNO} / r_{SNH}	Removal Rate for S_{NO} / S_{NH}	-	$gN.m^{-3}.d^{-1}$

2.2. Operational Data

The table 2 summarizes the conditions presented by the SBR process. They correspond to the mean wastewater composition of a real dairy cheese factory with a sole reactor operating in an IAEF configuration with carbon and nitrogen removal purposes.

Table 2: Plant Conditions

WW Composition & Flow Rate [SBR Inlet][*]	2000 mgCOD$_{Total}$/L; 140 mgN–NH$_4^+$/L; 20 mgN-NO$_3^-$/L & 800 m^3/d		
Cycle Time // Total Cycles by Day	12 h. // 2		
Reactor Sequencing for Each Cycle [Fixed]	Stage 1: Reaction Time 11 h.	Stage 2: Settling 0.5 h.	Stage 3: Draw 0.5 h.
Reactor Volume at Cycle Beginning	3000 m^3		
X_H & X_A at Cycle Beginning	2500 & 25 mgVolatile Suspended Solids(VSS)/L [experimental data]		
Sludge Retention Time	> 20 days		
DO Concentration	Not growth limiting (DO > 2 mgO$_2$/L)		

(*) Anaerobically pre-treated effluent

2.3. Constrained Dynamic MINLP Problem

The total aeration demand in the SBR cycle was assumed as the scalar performance index to be minimized. The batch reactor network size (integer variable), sequencing (binary variable: oxic/anoxic) and the stages duration (continuous variables), were assumed as the decision variables for the dynamic MINLP problem. These inputs allow constructing an optimal control law that can be easily implemented at industrial scale (switching times in the equipment). In addition to some physical path constraints, the environmental restrictions respect to the draw effluent composition defines the feasible region of the optimization problem. In this sense, the following inequalities terminal constraints must be satisfied: $COD_{Soluble}<60$ mg/L, $NH_4^+<5$ mgN/L, $NO_3^-<10$ mgN/L. The complete constrained dynamic MINLP problem to be solved is presented below:

$$\min_{u(t)} \quad J = \sum_k t_{Oxic,k}$$

$$s.t.:$$

$$\dot{c} = F(c, u) \qquad c(0) = c_0$$

$$S(c, u) \leq 0 \qquad T\big[c(t_f)\big] \leq 0$$

Where **J** is the objective function; $\mathbf{t_{Oxic,k}}$, the aeration time for the *k* stage in the SBR; **c**, the five-dimensional vector of states (COD, NH_4^+, NO_3^-, X_H, X_A) with known initial conditions $\mathbf{c_0}$ (60, 5, 10, 2500, 25); **u**, the m-dimensional vector of inputs (decision variables); **S**, the z-dimensional vector of path constraints; **T**, the three-dimensional vector of terminal constraints in accordance to the environmental regulations; **F**, the vector function who defines the ODE system (see *SBR Model and Kinetic Description*); and $\mathbf{t_f}$, the cycle time assumed fixed for this equipment (see Table 2). The function **u(t)** is parameterized in accordance to a piecewise constant control (Srinivasan et. al; 2003). The following three main inputs were chosen for such parameterization: **N**, the reaction network size (integer variable defining the number of O/A sub-sequences in the network); **I**, a binary variable who defines the first kind of reaction appearing in the process [oxic (**I = 1**); anoxic (**I = 0**)]; and $\mathbf{\tau_{First}}$, the duration of the stage located first at each O/A sub-sequence along the reaction network (continuous variable). To simplify the implementation of the optimal control law in the real equipment (PLC programming), every O/A sub-sequence in the process was assumed presents the same switching times. Thus, the objective function for this work can be represented as: **J = N.[I.τ_{First} + (1-I).(Reaction Time/N - τ_{First})]**, and the following path constraints were assumed for the inputs: **1 ≤ N ≤ 20; 0 < τ_{First} ≤ Reaction Time/N**; and **I.(Reaction Time/N - τ_{First}) + (1-I).τ_{First} ≤ 30 min.**. The last constraint must not be eluded in order to avoid the biomass settling during each anoxic stage (reactor not provided with an independent mixer). An adequate mixing in the system by short times once the aeration stopped (maximum ~ 30 minutes), can be achieved as a consequence of the N_2 generation (gas bubbles) during the denitrification process.

2.3.1. Deterministic Methods

Sequential shooting techniques/NLP Codes [Levenberg-Marquardt, Conjugate-Gradient, and Quasi-Newton (with steepest descent/linesearch algorithms)] (Srinivasan et. al; 2003, Biegler and Grossmann; 2004) were here employed to solve the optimal control law. It should be pointed out that the mixed-integer nature in the optimization problem was reached through the inclusion of additional constraints for **N** and **I** inputs. As a MINLP problem must be solved, the use of variational dynamic optimization methods (Pontryagin's Minimum Principle; Optimality Principle of Hamilton − Jacobi - Bellman; etc.) is not allowable (Srinivasan et. al; 2003, Biegler and Grossmann; 2004).

2.3.2. Stochastic Methods

Two kinds of PSO structures were here developed in order to calculate the optimal inputs: **mi-PSO** (mixed-integer PSO) and **r-PSO** (relaxed PSO). The difference resides in the way the mixed-integer nature in the optimization problem is accounted for. While **mi-PSO** restricts the input nature at each stage in the swarm movement, **r-PSO** permits a relaxed particle motion introducing the mixed-integer nature just when the objective function is computed. In order to meet the path and terminal constraints of the dynamic MINLP problem, the following augmented objective function is assumed for both PSO algorithms.

$$J = \begin{cases} N.\left[I.\tau_{First} + (1-I)\left(\dfrac{\text{Reaction Time}}{N} - \tau_{First}\right)\right] & \text{if } (N, I, \tau_{First}) \in \text{Feasible Region} \\ J_{Maximum} = \text{Reaction Time (Fixed Value)}_{(\text{see Table 2})} & \text{Otherwise} \end{cases}$$

3. Results and Discussion

3.1. Deterministic Optimization

Table 3 summarizes the optimal input values and the total aeration demand corresponding to four reaction networks obtained from different initial guesses and NLP codes. These results reveal the non convex behaviour presented in the Sequential Batch Reactor technology. Consequently, the application of deterministic techniques to optimize this kind of equipment will be only helpful if they are preceded by the use of an adequate global optimization method (heuristic, stochastic, etc.). In addition, if a powerful technique is employed to search for the global optimum [Genetic Algorithms, Scatter Search, Ant Colony Optimization, Particle Swarm Optimization, etc.], further deterministic contributions may not be necessary.

Table 3: Sequential Shooting Results

Shoot	NLP Code[*]	N	I	Reaction Network	τ_{First} (min)	Total Aeration Needs [J/J_{Max}] (%)
1		10	0	$AO(AO)_8AO$	30.0	54.5
2	Conjugate-Gradient	13	0	$AO(AO)_{11}AO$	30.0	40.9
3		16	1	$OA(OA)_{14}OA$	16.9	40.9
4		19	1	$OA(OA)_{17}OA$	14.1	40.6

[*] – Method that Optimize the Computational Cost.

3.2. Stochastic Optimization

As a constrained optimization problem must be solved, a large scale particle swarm should be generated in order to assure in the zero iteration, the random allocation of a minimal population sample (at least ~ 10 particles) inside the feasible region. As a result, and once the algorithm in course, the whole particle swarm would be gradually pushed inside the feasible region improving progressively their exploratory capabilities. Populations with 500 individuals, departing from repose and along 30 iterations were here used for both developed PSO algorithms. Table 4 shows the stochastic optimization results for **mi-PSO** and **r-PSO**.

Table 4: Particle Swarm Optimization Results (after 30 iterations)[*]

Method	N	I	Reaction Network	τ_{First} (min)	Total Aeration Needs [J/J_{Max}] (%)	Particles Inside Feasible Region [PIFR] (%)	Swarm Convergence Index [% of Population][**]
mi-PSO	15	0	$AO(AO)_{13}AO$	27.2	38.3	100	89.6
r-PSO	17	0	$AO(AO)_{15}AO$	23.9	38.4	100	98.6

[*] – Swarm Temporal Step (**Δt**) = 1; and Swarm Deadening Constant (**K**) = 2 for both PSO structures.
[**] – **Swarm Convergence Index (SCI) = 100.Total Converged Particles/Swarm Size**. It is assumed a Converged Particle when $|J_{Best,Particle} - J_{Leader,swarm}| < SCT.J_{Leader,swarm}$; *SCT* = Swarm Convergence Tolerance (assumed 0.1 in this table).

These results show that stochastic optimization improved the obtained results from deterministic techniques with a very reasonable computational cost (no further optimization improvement is needed). In addition, both PSO structures could allocate the complete swarm inside the feasible region (**PIFR = 100%**) demonstrating the enormous potential that this technique has. For identical computational efforts (after 30 iterations), the mixed-integer algorithm resulted slightly better than the relaxed one in order to optimize the constrained dynamic MINLP problem presented in this paper ($J_{miPSO} < J_{rPSO}$). However, **r-PSO** presented a higher swarm convergence index (**SCI**) than **mi-PSO** revealing very different natures in their exploratory capabilities. The **SCI** comparison strictly depends on the *SCT* value adopted for the calculations. Figure 2 shows a more exhaustive contrast between both PSO structures analyzing their complete response (**SCI & PIFR** indexes) along the course of the iterations. This data reveals very interesting additional results. Firstly, the higher **SCI** value for **r-PSO** is also observed during the whole course of the algorithm. Moreover, **mi-PSO** seems never reach a complete swarm convergence potential (**SCI = 100%**) because this performance index, even for different *SCT* values, always tend to a low asymptotic flat value as the iterations increases.
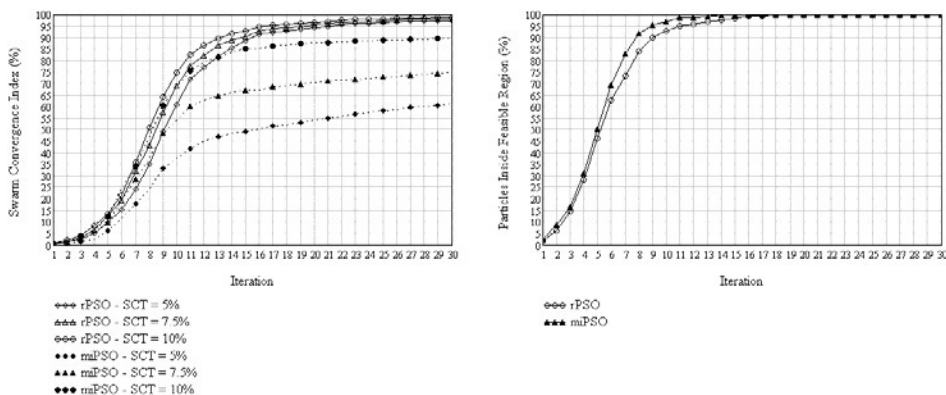
Figure 2: Comparison of **PSO** Indexes for Mixed-Integer and Relaxed Algorithms.

Finally, in order to reduce the computational cost for both structures, it can be observed that 15 – 20 iterations could be acceptable enough in order to get good optimization results (**PIFR** ~ **100%**; and **SCI** profiles already stabilized). In order to assure the most robust and reliable results for the MINLP problem presented in this work (high swarm convergence and good optimization efficiency), the assessment of both algorithms (**r-PSO** and **mi-PSO**) must be considered.

4. Conclusions

Optimal control laws obtained from deterministic and stochastic techniques were here developed for a real SBR process. New insights were obtained through the analysis of the energy management (aeration demand) in the equipment. Deterministic optimization results exposed a strong non convex behaviour for this system demonstrating the unavoidable need of stochastic contributions in order to search for the global optimum. The use of Particle Swarm Optimization allowed improving the obtained results from deterministic techniques with a very reasonable computational cost (no further optimization improvement was needed). In order to solve the constrained dynamic MINLP problem presented in this work and despite higher swarm convergences were obtained with a relaxed PSO formulation, better optimization results were achieved with a mixed-integer particle swarm structure, introducing the MI nature of the inputs at each stage in the algorithm. The assessment of both PSO formulations should be considered in order to guarantee the most robust and reliable results for this optimization problem. An optimal control law who minimizes the total aeration demand along the SBR operation can be easily programmed (switching times) in accordance to PSO results.

References

Artan, N.; Orhon, D. (2005). *Mechanism and Design of Sequencing Batch Reactors for Nutrient Removal.* Scientific and Technical Report No. 19, IWA Publishing.

Biegler, L.T. and Grossmann, I.E. (2004); Retrospective on Optimization. *Computers and Chemical Engineering*, 28, 1169–1192.

Coelho, M.A.Z.; Russo, C. and Araújo, O.Q.F. (2001); Optimization of a Sequencing Batch Reactor for Biological Nitrogen Removal. *Water Research,* 34(10), 2809-2817.

[1]Ferrari, A.; Biscaia Jr., E.C.; Melo, P.A. (2008); Sequential Batch Reactor and Plug Flow Reactor Network Comparison under Dynamic Conditions for Wastewater Treatment. *Proceedings of the 10th International Chemical and Biological Engineering Conference CHEMPOR 2008* (Braga – Portugal), 138 – 146.

[2]Ferrari, A.; Gutiérrez, S.; Benítez, A.; Canetti, R. (2008); Modelado Cinético y Desarrollo de Ley de Control para Reactores Batch Secuenciales Lácteos; *Memorias del IV Encuentro Regional de Ingeniería Química/2° Feria de la Industria Química y de Procesos EXPOQUIM 2008* (Montevideo – Uruguay). Paper in Spanish.

Kennedy, J. and Eberhart, R. (1995); Particle Swarm Optimization. *Proceedings of IEEE International Conference on Neural Networks* (Perth - Australia), 1942-1948.

Orhon, D.; Artan, N. (1994). *Modelling of Activated Sludge Systems.* Technomic Publishing Company Inc.

Souza, S.M.; Araújo, O.Q.F. and Coelho, M.A.Z. (2008); Model-based optimization of a sequencing batch reactor for biological nitrogen removal. *Bioresource Technology*, 99(8), 3213-3223.

Srinivasan, B.; Palanki, S. and Bonvin, D. (2003); Dynamic optimization of batch processes. I. Characterization of the nominal solution. *Computers and Chemical Engineering*, 27, 1–26.

Addressing Long-Term Biorestoration in Eutrophic Lakes as an Optimal Control Problem, Under Different Scenarios

Vanina Estrada [a], Elisa R. Parodi [b], M. Soledad Diaz [a]

[a]*Planta Piloto de Ingeniería Química (PLAPIQUI), Universidad Nacional del Sur-*
CONICET, *Camino La carrindanga Km 7, Bahía Blanca 8000, Argentina,*
sdiaz@plapiqui.edu.ar
[b]*Instituto Argentino de Oceanografía (IADO), Universidad Nacional del Sur-*
CONICET, *Camino La carrindanga Km 7, Bahía Blanca 8000, Argentina,*

Abstract

In this work we propose an ecological water quality model for a reservoir, based on first principles, to study the response of main phytoplankton groups (algae) to temperature rise due to global warming and different nutrient (phosphate) loading scenarios in the middle and long term. An optimal control problem is formulated to determine optimal combination of restoration strategies, including nutrient loading reduction through inflows remediation within a wetland and in-lake biorestoration techniques. The dynamic optimization problem is formulated within a simultaneous approach.

Keywords: Eutrophication, Optimal control, Dynamic optimization, Climate change, Nutrient loading.

1. Introduction

Lakes and reservoirs over-enrichment by plant nutrient due to urban, agricultural, and industrial development has accelerated the natural process that takes place in water bodies and is referred to as eutrophication. This problem is associated to excessive growth of phytoplankton (algal blooms), especially cyanobacteria, which may produce hepato- and neurotoxins that can severely compromise human and animal health. Cyanobacterial blooms take place during warm months because their maximum growth rates are attained at temperatures above 25 ºC. Temperature is a key parameter in most biological systems and it is suggested that global warming could directly affect several processes in freshwater bodies, e. g. growth, respiration and death rate of phytoplankton, nutrients release from lake sediments (Komatsu et al., 2007). Since cyanobacteria grow on high temperatures, climate change may encourage their growth and they could increasingly dominate the community. Biorestoration in water bodies is a technique based on trophic chain theory and has been one of the most applied to control phytoplankton growth in lakes and reservoirs. The basic idea is to keep a high grazing pressure on the phytoplankton community by the herbivore zooplankton by performing zooplanktivorous fish removal (Sondegaard et al., 2007). However, the middle and long term effect of restoration is not well documented and has had different experimental results depending on the specific water bodies reported in the literature (Søndergaard et al., 1997). A few papers have addressed simulation of the evolution of water bodies under biomanipulation (Sagehashi et al., 2001, Gurkan et al., 2006).

Estrada et al. (2008) have formulated an optimal control problem for algae growth control along a one year horizon, based on a hybrid ecological water model.

In this work, we propose a hybrid eutrophication model within an optimal control problem to determine restoration policies for water quality improvement, taking into account two climate change scenarios for Argentina up to 2013 and two scenarios of nutrient loading, with special attention to the response of cyanobacteria. The hybrid biogeochemical model has been formulated within a simultaneous dynamic optimization approach (Kameswaran and Biegler, 2006) whose solution provides limiting nutrient inflow profiles to the lake, as well as in-lake bio-manipulation profiles.

2. Process description and forecast scenarios

Paso de las Piedras Reservoir is located in the south of Buenos Aires Province (Argentina) at 38° 22′ S and 61° 12′ W. It has 60 km of coastline perimeter, 36 km^2 surface area and a mean depth of 8.2 m. It supplies drinking water to two cities (population around 450,000) and to a petrochemical complex. The reservoir has two tributaries, Sauce Grande River and El Divisorio Stream (6.69 and 0.64 m^3/s mean flowrates, respectively), which run through an important agricultural area. High nutrient concentration in the water body is due to loading from tributaries, as the reservoir is within a protected area (Paso de las Piedras Provincial Park). Its current trophic state is eutrophic and there is an urgent need for restoration. Weekly data on concentrations, temperatures, inflows, outflows, solar radiation etc, collected throughout an entire year in 2004 at four sampling stations in the reservoir, have made possible the development of a first principles hybrid water quality model (Estrada et al., 2009). Inlet variable profiles for the model have been obtained as time correlations for temperature, solar radiation, inflows flowrate and concentrations, based on collected data. In the present work, we study the effect of climate warming for the next five years upon both main phytoplankton and nutrient concentration in Paso de las Piedras reservoir and restoration policies. We have selected two scenarios that represent the lowest and highest temperature forecasts for Argentina, based on greenhouse gases emission models (http://www.cru.uea.ac.uk/~mikeh/research/wwf.argent.pdf). Furthermore, regarding nutrient loading, phosphate concentration profiles in the two tributaries have been reduced to a half and increased twice their current annual mean value (Sauce Grande River 18.53 mg/m^3 and El Divisorio Stream 90.26 mg/m^3) to study its influence on remediation strategies.

3. Water quality modeling

The proposed water quality ecological model is based on first principles, with parameters that have been tuned with collected data from the specific reservoir under study (Estrada et al., 2009). It includes dynamic mass balances for three phytoplankton groups (cyanobacteria, diatoms and chlorophyta) and main nutrients. Input variable profiles are represented as sinusoidal functions in time. Detailed description of the model equations are given in Estrada et al. (2009). The main simplifying assumption is the consideration of horizontally averaged concentrations. This assumption has been introduced in the literature in most mechanistic eutrophication models for lakes and reservoirs (Arhonditsis et al., 2006, Zhao et al., 2008). To decide the inclusion of this assumption, we analyzed collected data at the four existing sampling stations. The small relative differences found (below 10%) justify the assumption of horizontally averaged concentrations in our model. Additional simplifying assumptions in the present model include constant transversal area in the reservoir and constant water density. The

resulting partial differential equations system (considering concentration gradients along the water column) has been discretized into two water layers and main equations for each layer are given below.

Total mass balance

$$\frac{dh_T}{dt} = \frac{1}{A} \sum_{k=1}^{NIN} QIN_k - \frac{1}{A} \sum_{m=1}^{NOUT} QOUT_m + Q_{rain} - Q_{evap} \tag{1}$$

Component mass balances for horizontal layers (U: upper layer, L: lower layer: j: Cyanobacteria, Diatoms, Chlorophyta, Nitrate, Ammonium, Organic Nitrogen, Phosphate, Organic Phosphorus, Biochemical Demand of Oxygen, Dissolved Oxygen)

Upper layer

$$\frac{dC_{Uj}}{dt} = \sum_{k=1}^{NIN} \frac{QIN_{U,k}}{V_U} CIN_{Ujk} - \sum_{m=1}^{NOUT} \frac{QOUT_U}{V_U} C_{Uj} + r_{Uj} - \frac{kdA}{\Delta h_U h_U}(C_{Uj} - C_{Lj}) - \frac{C_{Uj}}{h_U}\frac{dh_U}{dt} \tag{2}$$

Lower layer

$$\frac{dC_{Lj}}{dt} = \sum_{m=1}^{NOUT} \frac{QOUT_L}{V_L} C_{Lj} + r_{Lj} + \frac{kdA}{\Delta h_L h_L}(C_{Lj} - C_{U,j}) - \frac{C_{Lj}}{h_L}\frac{dh_L}{dt} \tag{3}$$

Phytoplankton: Rate equations for phytoplankton groups take into account production and losses due to respiration, natural death, settling and grazing. The growth rate of the three phytoplankton groups is calculated with a multiplicative model as function of solar radiation, water temperature and nutrients availability, as follows:

$$R_{ij,growth} = k_{i,growth} f(T)_{ij} f(I)_{ij} f(N)_{ij} C_{ij} \tag{4}$$

$$f(T)_{ij} = -\frac{(T_j - T_{opt_i})^2}{T^2_{opt_i}} + 1 \quad f(I)_{ij} = \frac{I_{oi}}{I_{opt_j}} exp\left(1 - \frac{I_{oi}}{I_{opt_j}}\right) \quad f(N)_{ij} = \frac{C_{PO_{4j}}}{C_{PO_{4j}} + K_{P_i}} \tag{5}\,(6)\,(7)$$

Phytoplankton respiration, natural death, settling rates to the sediment and herbivorous zooplankton grazing rate are calculated as:

$$R_{ij,resp} = k_{j,resp}\theta_r^{(T-20)}C_{ij} , \quad i = UL, LL; j = cyano, diatom, chlorophyte \tag{8}$$

$$R_{ij,death} = k_{j,death}\theta_m^{(T-20)}C_{ij} , \quad i = UL, LL; j = cyano, diatom, chlorophyte \tag{9}$$

$$R_{ij,settling} = k_{j,settling}\frac{C_{ij}}{h_i} , \qquad i = UL, LL; j = cyano, diatom, chlorophyte \tag{10}$$

$$R_{ij,graz} = k_{j,graz}\frac{C_{ij}}{C_{ij} + K_{graz}}C_{i,Zoo} , \, i = UL, LL; j = cyano, diatom, chlorophyte \tag{11}$$

Nutrient cycles (phosphorus and nitrogen): State variables describing phosphorus cycle are phosphate and organic phosphorus. Phosphorus is uptaken by phytoplankton in as phosphate. As phytoplankton biomass is composed of carbon, nitrogen and phosphorus, upon death, phytoplankton increases both phosphate and organic phosphorus and organic nitrogen pool. Organic phosphorus reacts to phosphate, which is then available for phytoplankton generation again. Three state variables describe nitrogen cycle: ammonium, nitrate and organic nitrogen. Phytoplankters uptake both ammonium and nitrate for growth. Organic nitrogen is hydrolyzed at to ammonium, which is in turn oxidized to nitrate. For the lower layer, the model includes a term of ammonium release from sediments. Nitrate can be reduced to molecular nitrogen by denitrification. Detailed rate equations are given in Estrada et al. (2009).

4. Biorestoration and nutrient loading reduction in the middle and long term as an optimal control problem

The DAE system, partially described in the previous section, has been formulated within a dynamic optimization problem whose objective function is the minimization of the weighted sum of the offset between cyanobacteria concentration in the upper water layer and a desired value of 0.10 mg/l and the offset between phosphate concentration and a desired value of 0.02mg/l (both desired values below eutrophication limits)

$\left(\min\left[0.7 * \int_0^{tf} \left(\sum_{j=phyto} C_j(t), -0.10 \right)^2 dt + 0.3 * \int_0^{tf} \left(C_{PO4}(t) - 0.02 \right)^2 dt \right] \right)$ throughout a time horizon of five

years. Not only nutrient loading reduction is required for algae growth control, but in-lake restoration techniques are additionally required due to the complex nutrient recycles previously described (Sondegaard et al., 2007). The proposed model has two time dependent degrees of freedom to represent these strategies: flowrate of tributary that is derived to a nearby wetland for remediation (*Fwetland*) and zooplankton concentration ($c_{U,zoo}$) within the reservoir. The first optimization variable represents handling of nutrient loading to the reservoir through the fraction of inflows that is derived to an existing wetland for nutrient absorption. The wetland size and retention capacity have been included in the model to impose bounds on the maximum derivation flowrate (50% of El Divisorio Stream flowrate) and nutrient concentration reduction (50% global retention of phosphate in the wetland). The second optimization variable ($c_{U,zoo}$) stands for one of the most efficient in-lake restoration strategies, based on the food chain concept: removal of zooplanktivorous fish to increase the possible top-down control of zooplankton on the phytoplankton (Sondegaard et al., 2007). Zooplankton concentration along the time horizon is considered a measure of fish removal rate by including fish biomass data from the specific lake.

The dynamic optimization problem is formulated within a simultaneous dynamic optimization framework by transforming it into a large-scale nonlinear program (NLP) by collocation on finite elements. We have used an Interior point method within program IPOPT (Biegler et al., 2000) with reduced SQP techniques for solving the large-scale NLP. Extensions of this approach are described in Kameswaran and Biegler (2006).

5. Case study: Control of algae growth in Paso de las Piedras under different nutrient scenarios and global warming

As a first step, we have solved the DAE system representing water quality evolution along a time horizon of five years, on a daily basis, for the three temperature profiles shown in Fig. 1, corresponding to current, lower and upper forecasts calculated with green house emission models for Argentina. The DAE model has also been solved for different nutrient (phosphate) scenarios, including current situation and half and double its current inflows, shown in Fig. 2. Simulation results show that there is an increase in cyanobacteria biomass (mgC/l) annual mean of 0.31% in the first year, 8.26% in the second year and 10.4% in the last three years when phosphate loading is increased twice its current profile, while there is no change in the annual mean of cyanobacteria concentration when the model was run with half phosphate inflows concentration ($\sim 10^{-5}$%) and with the different temperature scenarios (less than 0.01%). In this sense, numerical results show that the change in water temperature predicted for the next five years has much less significant effect on annual cyanobacteria biomass than the increase in phosphate loading.

We have solved the optimal control problem described in Section 4 for the different scenarios. The dynamic optimization problem for two water layers has 23 differential equations and 60 algebraic ones. Within a simultaneous optimization approach, it has been discretized in time with 100 finite elements and two collocation points rendering a nonlinear problem with 19523 variables, which has been solved with program IPOPT (Biegler et al., 2002) that implements an Interior Point method with reduced Successive Quadratic Programming. The optimization variables are the fraction of inlet stream that is derived to the wetland and the concentration of zooplankton in the lake to control phytoplankton growth, along the five years time horizon. Numerical results show that tributary deviation through a wetland for nutrient loading reduction is required throughout the entire time horizon, at its maximum allowed flowrate (50% of the inflows, with 50% nutrient retention) when the second term (minimization of phosphate offset) is included in the objective function. Figures 3 and 4 show cyanobacteria concentration profiles before and after biomanipulation, as well as zooplankton concentration profiles for the model run in normal conditions and for double phosphate inflows concentration. Nutrient loading reduction requires even longer time horizons to show noticeable effects on concentration (Fig. 5) due to internal nutrient recycle whereas increases in phosphorus loading has an immediately effect on cyanobacteria concentration.
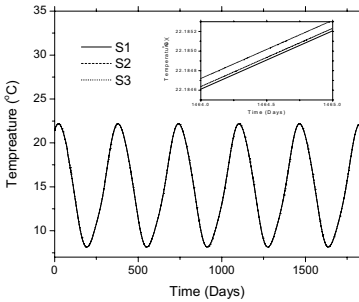
Figure 1. Temperature scenarios. S1: current temperature, S2: lower predicted temperature, S3: upper predicted temp.
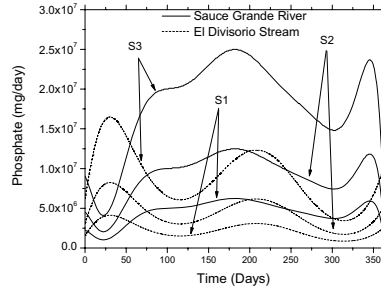
Figure 2. Phosphate inflows scenarios. S1: current PO_4 inflows, S2: half PO_4 inflows conc., S3: twice PO_4 inflows concentration
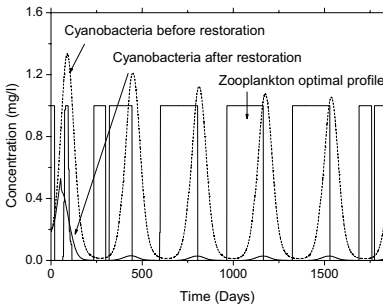
Figure 3. Comparison between cyanobacteria concentration profiles before and after restoration and optimal profile for zooplankton conc. with a current conditions of PO_4 inflows.
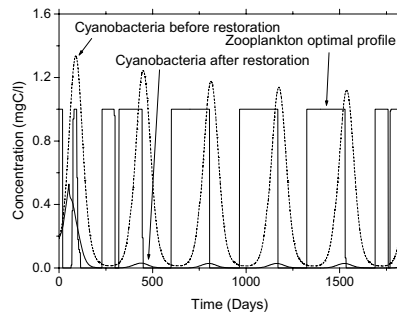
Figure 4. Comparison between cyanobacteria concentration profiles before and after restoration and optimal profile for zooplankton conc. with a twice PO_4 inflows concentration.
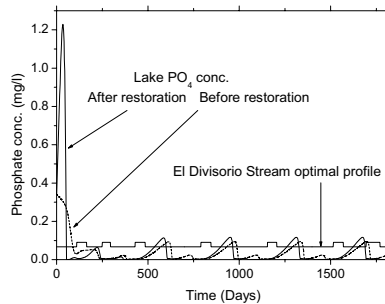
Figure 5. Comparison between phosphate concentration profiles in Paso de las Piedras reservoir before and after restoration and optimal PO_4 conc. profile from El Divisorio Stream.

6. Conclusions

Advanced dynamic optimization techniques integrated to first principles hybrid models for water bodies have given useful insights on the determination of biorestoration policies. The formulation of an optimal control problem for a long term horizon has provided quantitative information on the combined application of nutrient loading reduction strategies by partial remediation of inflows within an existing wetland and the biomanipulation technique of zooplanktivorous fish removal for top-down control of zooplankton on phytoplankton. Numerical results have also shown that global warming along the next five years has no noticeable effect on algal growth. On the other hand, increases in nutrient loading, mainly due to nonpoint sources (agricultural activities) will certainly affect lake quality if no restoration action is taken.

References

G. B. Arhonditsis, C. A. Stow, L. J. Steinberg, M. A. Kenny, R. C. Lathrop, S. J. McBride and K. H. Reckhow, 2006, Exploring ecological patterns with estructural equation modeling and Bayesian analysis, Ecol. Model., 192, 385-409.

L. T. Biegler, Cervantes, A. and Waechter, A., 2002, Advances in simultaneous strategies for dynamic process optimization, Chem. Eng. Sci., 57, 575-593.

V. Estrada, E. R. Parodi and M. S. Diaz, 2008, Addresing the control problem of algae growth in water reservoirs with advanced dynamics optimization approaches, FOCAPO 2008, USA.

V. Estrada, E. R. Parodi and M. S. Diaz, 2009, Determination of biogeochemical parameters in eutrophication models with simultaneous dynamic optimization approaches, Comp. & Chem. Eng., in press.

Z. Gurkan, J. Zhang, and S. E. Jørgensen, 2006. Development of a structurally dynamic model for forecasting the effects of restoration of Lake Fure, Denmark, Ecol. Model., 197, 89-102.

S. Kameswaran and L. T. Biegler, Simultaneous dynamic optimization strategies: Recent advances and challenges, 2006, Comp. & Chem. Eng., 30, 1560-1575.

E. Komatsu, T. Fukushima and H. Harasawa, 2007, a modelin approach to forecast the effect of long-term climate change on lake water quality, Ecol. Model., 209, 351-366.

M. Sagehashi, A. Sakoda and M. Suzuki, M., 2001, A mathematical model of a shallow and eutrophic lake (The Keszthel Basin, Lake Balaton) and simulation of restorative manipulations, Wat. Res., 35, 1675-1686.

M. Søndergaard, E. Jeppesen, T. L. Lauridsen, C. Skov, E. H. van Nes, R. Roijackers, E. Lammens, and R. Portielje, 2007, Lake restoration: successes, failures and long-term effects. J. Appl. Ecol., 44, 1095–1105.

J. Zhao, M. Ramin, V. Cheng and G. B. Arhonditsis, 2008, Plankton community patterns across a trophic gradient: The role of zooplankton functional groups, Ecol. Model. 213, 417-436.

Design of Stable Large-Scale Metabolic Networks

Jimena Di Maggio, Aníbal M. Blanco, J. Alberto Bandoni, M. Soledad Diaz

Planta Piloto de Ingeniería Química (PLAPIQUI),
Universidad Nacional del Sur-CONICET,
Camino La Carrindanga Km 7, Bahía Blanca 8000, Argentina

Abstract

In this work we propose an eigenvalue optimization approach to ensure steady state stability of the Embden-Meyerhof-Parnas pathway, the pentose-phosphate pathway and the phosphotransferase system of *Escherichia coli*. The model consists of eighteen differential equations that represent dynamic mass balances for extracellular glucose and intracellular metabolites and thirty kinetic rate expressions. The nonlinear optimization problem including stability constraints has been solved with reduced space Successive Quadratic Programming techniques within program IPOPT (Waechter and Biegler et al., 2006). Numerical results provide useful insights on the stability properties of the studied kinetic model.

Keywords: Metabolic networks, Stability, Eigenvalue optimization.

1. Introduction

Metabolic networks design can be formulated as an optimization problem aimed at optimizing a given objective, for example the production of a certain metabolite, subject to mass balance equations that represent the network. Kinetic models allow the analysis of stability of the predicted states, which is of fundamental importance because biological systems may exhibit monotonic stable states, bistable switching threshold phenomena, oscillations and chaotic behavior. Due to nonlinear kinetics of the biochemical reactions and their coupling through common metabolites, biological systems may undergo drastic changes in their qualitative behavior when a variation on the enzyme level occurs. If no stability constraints are included in the formulation, the optimal operating point might be unstable, making the metabolic network vulnerable to external disturbances. In other words, in spite of the presence of modest disturbances an unstable network will reach physiological constrains and collapse. Several authors have addressed the analysis if biological systems of small to moderate size (Hatzimanikatis and Bailey, 1997, Haddad and Chellaboina, 2005). The design-for-stability problem, an important sub problem of the general design-for-operability problem, has also motivated many contributions from the process systems engineering community. Different strategies have been proposed to include stability considerations within the design problem (Chang and Sahinidis, 2004; Blanco et al., 2004).

In this work we propose an eigenvalue optimization approach (Blanco and Bandoni, 2007) to ensure steady state stability of the glycolysis, the pentose-phosphate pathway and the phosphotransferase system of *Escherichia coli* K-12 W3110 (Chassagnole et al., 2002). The nonlinear optimization problem, corresponding to steady state equations and stability constraints, has been solved with reduced space Successive Quadratic Programming techniques within program IPOPT (Waechter and Biegler, 2006). Optimization results provide an improved metabolic network for the maximization of

serine production in *Escherichia coli*, within the steady state stable region, based on a detailed kinetic model.

2. Optimization model description

In this work, we have studied an adaptation of the dynamic model for the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway and the phosphotransferase system, as shown in Fig. 1, of *Escherichia coli* K-12 W3110 (Chassagnole et al., 2002). It comprises eighteen differential equations that represent dynamic mass balances of extracellular glucose and intracellular metabolites, thirty kinetic rate expressions and it involves one hundred and sixteen parameters. Most influential parameters have been determined through previous work on global sensitivity analysis of the proposed model (Di Maggio et al., 2008a,b).

Equations (1) to (6) correspond to main mass balances on metabolites involved in the metabolic network shown in Fig. 1. A detailed description of the remaining twelve balances is given in Chassagnole et al. (2002).

$$\frac{dC_{glc}^{ext}}{dt} = D(C_{glc}^{a\lim} - C_{glc}^{ext}) + f^{pulso} - \frac{C_X r_{PTS}}{\rho_X} \tag{1}$$

$$\frac{dC_{g6p}}{dt} = r_{PTS} - r_{PGI} - r_{G6PDH} - r_{PGM} - \mu C_{g6p} \tag{2}$$

$$\frac{dC_{f6p}}{dt} = r_{PGI} - r_{PFK} + r_{TKb} + r_{TA} - 2r_{MurSynth} - \mu C_{f6p} \tag{3}$$

$$\frac{dC_{gap}}{dt} = r_{ALDO} + r_{TIS} - r_{GAPDH} + r_{TKa} + r_{TKb} - r_{TA} + r_{TrpSynth} - \mu C_{gap} \tag{4}$$

$$\frac{dC_{pep}}{dt} = r_{ENO} - r_{PK} - r_{PTS} - r_{PEPCxylase} - r_{DAHPS} - \mu C_{pep} \tag{5}$$

$$\frac{dC_{pyr}}{dt} = r_{PK} + r_{PTS} - r_{PDH} + r_{MetSynth} + r_{TrpSynth} - \mu C_{pyr} \tag{6}$$

Equations 7 to 13 show kinetic expressions for phosphotransferase system, glucose-6-phosphate isomerase, 6-phosphogluconate dehydrogenase, phosphoglycerate kinase and phosphoglycerate mutase, the Serine synthesis pathway and the Chorismate and Mureine synthesis pathway, respectively. The remaining rate equations involved in the present metabolic network can be found in Chassagnole et al. (2002).

$$r_{PTS} = \frac{r_{PTS}^{\max} c_{glc}^{extracelul\ lar} \frac{c_{pep}}{c_{pyr}}}{\left(K_{PTS,a1} + K_{PTS,a2}\frac{c_{pep}}{c_{pyr}} + K_{PTS,a3}c_{glc}^{extracelul\ lar} + c_{glc}^{extracelul\ lar}\frac{c_{pep}}{c_{pyr}}\right)\left(1 + \frac{c_{g6p}^{N_{PTS,g6p}}}{K_{PTS,g6p}}\right)} \tag{7}$$

$$r_{PGI} = \frac{r_{PGI}^{\max}\left(C_{g6p} - \frac{C_{f6p}}{K_{PGI,eq}}\right)}{K_{PGI,g6p}\left(1 + \frac{C_{f6p}}{K_{PGI,f6p}\left(1 + \frac{C_{6pg}}{K_{PGI,f6p,6pginh}}\right)} + \frac{C_{6pg}}{K_{PGI,g6p,6pginh}}\right) + C_{g6p}} \tag{8}$$

$$r_{PGDH} = \frac{r_{PGDH}^{\max} c_{6pg} c_{nadp}}{\left(c_{6pg} + K_{PGDH,6pg}\right)\left(c_{nadp} + K_{PGDH,nadp}\left(1 + \frac{c_{nadph}}{K_{PGDH,nadph,inh}}\right)\left(1 + \frac{c_{ATP}}{K_{PGDH,ATP,g6pinh}}\right)\right)} \tag{9}$$

$$r_{PGK} = \frac{r_{PGK}^{\max}\left(C_{adp}C_{pgp} - \frac{C_{atp}C_{3pg}}{K_{PGK,eq}}\right)}{\left(K_{PGK,adp}\left(1 + \frac{C_{atp}}{K_{PGK,atp}}\right) + C_{adp}\right)\left(K_{PGK,pgp}\left(1 + \frac{C_{3pg}}{K_{PGK,3pg}}\right) + C_{pgp}\right)} \tag{10}$$

$$r_{PGluMu} = \frac{r_{PGluMu}^{max}\left(C_{3pg} - \dfrac{C_{2pg}}{K_{PGluMu,eq}}\right)}{K_{PGluMu,3pg}\left(1 + \dfrac{C_{2pg}}{K_{PGluMu,2pg}}\right) + C_{3pg}} \tag{11}$$

$$r_{SerSynth} = \frac{r_{SerSynth}^{max}C_{3pg}}{K_{SerSynth,3pg} + C_{3pg}} \tag{12}$$

$$r_{Synth1} = \frac{r_{Synth1}^{max}C_{pep}}{K_{Synth1,pep} + C_{pep}} \tag{13}$$
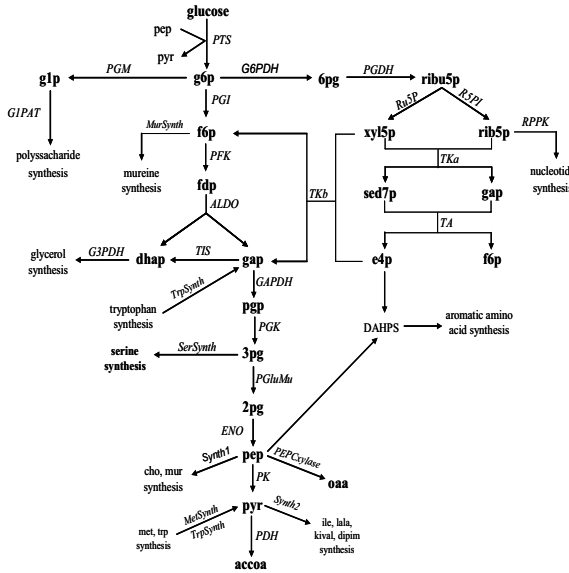
Figure 1. Metabolic network of glycolysis and pentose-phosphate pathway in *Escherichia coli*.

The maximum reaction rates, r_i^{max}, are associated to the corresponding enzyme concentration, so they could be tuned to maximize or minimize the production of a given metabolite. We have formulated an optimization problem in which the maximum reaction rates for certain pathways are the degrees of freedom. As changes in these parameters imply that the levels of enzymes will be modified, the total proteins concentration will vary and physiological changes that are not taken account in the dynamic model could be appear. To account for processes such as redistribution of limited mRNA contents and homeostasis (Mauch et al., 2001, Nikolaev et al., 2005), we have included equations (14) and (15). Eqn. (14) imposes that an increase in certain enzyme levels is compensated by a decrease in the remaining ones. Eqns. (15) ensure that gene expression rates of non-modulated enzymes be maintained at ratios equal to the ones at the reference steady state.

$$\frac{1}{M}\sum_{j=1}^{M}\frac{r_j^{max}}{r_j^{max,0}} = 1 \tag{14}$$

$$\frac{r_{j_1'}^{max}}{r_{j_1'}^{max,0}} = ... = \frac{r_{j_K'}^{max}}{r_{j_K'}^{max,0}} = \gamma \tag{15}$$

where $r_j^{max,0}$ is the maximum reaction rate at the reference state, M is the number of enzymes in the metabolic network, $j_1'...j_K'$ are the indices of non-modulated enzymes,

$K = M-L$ and L is the number of modulated enzymes (corresponding to maximum reaction rates being optimization variables).

Equations (14) and (15) can be re-written as (16) and (17), respectively, to be included in the optimization problem formulation.

$$\frac{r_{j_1}^{max}}{r_{j_1}^{max,0}} + ... + \frac{r_{j_L}^{max}}{r_{j_L}^{max,0}} + K\gamma = M \tag{16}$$

$$r_{j_S}^{max} = \gamma r_{j_S}^{max,0} \qquad , S = 1,...,K \tag{17}$$

3. Optimization under stability constraints

In order to assess asymptotic stability of dynamic systems, eigenvalue analysis is usually performed. For an asymptotically stable equilibrium point, the eigenvalues of the dynamic system Jacobian matrix lie on the left half of the complex plane. In an eigenvalue optimization problem (Blanco and Bandoni, 2007), the real parts of the eigenvalues of the Jacobian matrix of the dynamic system under study are forced to be strictly negative, ensuring that way asymptotic stability of the resulting equilibrium point. The eigenvalue optimization problem can be stated as follows:

$$\min_{y} \Phi(y)$$

s.t
$$A^T(y)P + PA(y) + I = 0$$
$$\det(P_i^{-1}) \geq \xi \qquad i=1,...,n$$
$$\xi > 0 \tag{18}$$
$$h(y) = 0$$
$$g(y) \leq 0$$
$$y \in Y$$

where y represents the optimization variables which comprises both, design and operating variables. $\Phi(y)$ is the objective function, $h(y)$ is the set of equality constraints and $g(y)$ is the set of inequality constraints. $A(y)$ is the Jacobian matrix of the dynamic system and P is a real symmetric matrix defined in (18) through the Lyapunov equation. Det(P_i^{-1}) stands for determinants of the principal minors of the inverse of P, I is the identity matrix and ξ a user defined positive constant. For details on the derivation and solution strategy of (18) see Blanco and Bandoni, (2007).

4. Discussion of results

In this study, and based on previous work on global sensitivity analysis on the dynamic metabolic network to main parameters (Di Maggio et al., 2008a,b), we have selected four maximum reaction rates as design variables, corresponding to serine synthesis ($r_{SerSynth}^{max}$), glucose-6-phosphate isomerase (r_{PGI}^{max}), phosphotransferase system (r_{PTS}^{max}) and Chorismate and Mureine synthesis (r_{Synth1}^{max}), respectively. The design problem to maximize serine production ($\Phi(y)=r_{SerSynth}$), for the steady state kinetic model of the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway and the phosphotransferase system comprises mass balances for eighteen metabolites and thirty rate equations associated to thirty enzyme levels (partially shown as Eqns. (7) to (13), as well as twenty seven equations ($K+1$) standing for Eqns. (16) and (17).

When formulating the design problem under stability constraints, Eqn. (18), additional equality constraints (eighteen) and inequalities (eighteen) are included in the optimization problem formulation, standing for Lyapunov's equation and nonnegativity on the determinants of the principal minors of P, to ensure its positive definiteness,

respectively. Nonlinear optimization models have been implemented in a Fortran 90 environment and they have been solved with reduced space Successive Quadratic Programming techniques within program IPOPT (Waechter and Biegler, 2006). The design problem has been first solved without stability constraints (Case 1) and it has been extended to include the proposed equations to ensure a stable optimal network design (Case 2).

Table 1. Main optimization variable values.

Optimization variables	Nominal values	Case 1	Case 2
C_{glcext} (mM)	1,7222	0,9946	0,2599
C_{g6p} (mM)	3,4800	5,1859	4,2877
C_{f6p} (mM)	0,6000	0,8941	0,7393
C_{fdp} (mM)	0.2720	0,2211	0,3280
C_{gap} (mM)	0,2180	0,1818	0,2349
C_{dhap} (mM)	0,1670	0,1485	0,1838
C_{pgp} (mM)	0,0080	0,0035	0,0059
C_{3pg} (mM)	2,1300	0,8866	1,5158
C_{2pg} (mM)	0,3990	0,1656	0,2837
C_{pep} (mM)	2,6700	1,0491	1,7945
C_{pyr} (mM)	2,6700	3,1879	2,9868
C_{6pg} (mM)	0,8080	0,9927	0,8572
C_{ribu5p} (mM)	0,1110	0,1475	0,1385
C_{xyl5p} (mM)	0,1380	0,1713	0,1637
C_{sed7p} (mM)	0,2760	0,5147	0,3624
C_{rib5p} (mM)	0,3980	0,5089	0,4823
C_{e4p} (mM)	0,0980	0,1005	0,1110
C_{g1p} (mM)	0,6530	0,9661	0,7910
r_{PGI}^{max} (mM/sec)	495,870	442,316	495,842
$r_{SerSynth}^{max}$ (mM/sec)	0,0203	0,2000	0,0985
r_{Synth1}^{max} (mM/sec)	0,0148	0.0000	0.0000
Serine Production (mM/sec)	0,0138	0,0939	0,0594

Case 1: *Optimization for maximization of Serine production*
Numerical results show that serine production could be increased from 0,01381 mM/sec in a reference steady state (experimental) to 0,09396 mM/sec when the maximum serine synthesis reaction rate ($r_{SerSynth}^{max}$) is at its upper bound (=10* $r_{SerSynth}^{max,0}$). Main variable values are shown in the third column of Table 1. The eigenvalues of matrix $A(y)$ show that the metabolic network is stable, being the real part of the largest eigenvalue -3.7E-4. While stable, the system is close to critical stability and small changes in some parameters could easily lead to an unstable equilibrium. Taking into account the difficulty to modulate enzyme levels with precision, this implies that a relative deviation on the enzyme levels from their desired value may produce unstable behavior in the system. Therefore, the inclusion of stability constraints in the design problem becomes necessary.

Case 2: *Optimization for maximization of Serine production under stability constraints*
In order to modify the spectrum of matrix $A(y)$ different trials with parameter ξ were performed. For ξ= 1E-4, the largest eigenvalue in real part is -1.5E-3. In this case, maximum serine production can be increased 330% from its reference steady state value (0,01381 mM/sec) to 0,05936 mM/sec. It can be noted that some robustness of the

network regarding dynamic stability was achieved at the expense of a worsening in the adopted objective function. Furthermore, serine synthesis reaction rate ($r_{SerSynth}{}^{max}$) is not at its upper bound, as constraints on the determinants of the inverse of P become active. The fourth column in Table 1 shows main variable values in this case. In both Case 1 and 2, optimal values for the maximization of Serine production implies $r_{Synth1}{}^{max} = 0$. This corresponds to deletion of Chorismate and Mureine synthesis pathway.

5. Conclusions

The proposed approach in the formulation of a design problem for a large-scale kinetic model for a metabolic network including stability constraints, allows the determination of an improved network for serine production by ensuring its stability at the design level.

6. Acknowledgment

References

A.M. Blanco and J.A. Bandoni, 2007, Eigenvalue Optimization Based Formulations for Dynamics and Control Problems, Chemical Engineering Process, 46, 1192-1199

A.M. Blanco, J.A. Bandoni, L. Biegler, 2004, Re-design of the Tennessee Eastman Challenge Process: An Eigenvalue Optimization Approach, Proceedings FOCAPD 2004, Princeton, 517-520

Y. Chang, N. Sahinidis, 2004, Optimization of metabolic pathways under stability considerations, Computers and Chemical Engineering, 29, 467-479

C. Chassagnole, N. Noisommit-Rizzi, J. Schmid, K. Mauch, M. Reuss, 2002, Dynamic Modeling of the Central Carbon Metabolism of *Escherichia coli*, Biotechnology and Bioengineering, 79, 1, 53-73

J. Di Maggio, J.C. Diaz Ricci, M.S. Diaz, 2008a, Global Sensitivity Analysis: Estimation of Sensitivity Indices in Metabolic Network Dynamic Models, submitted to Metabolic Engineering.

J. Di Maggio, J.C. Diaz Ricci, M.S. Diaz, 2008b, Ranking of most influential kinetic parameters in metabolic networks through Global Sensitivity Analysis, Engineering Conferences International-Metabolic Engineering VII, September 14-19, Puerto Vallarta, Mexico

W. Haddad and V. Chellaboina, 2005, Stability and dissipativity theory for nonnegative dynamical systems: a unified analysis framework for biological and physiological systems, Nonlinear Analysis-Real World Applications, 6, 33-65

V. Hatzimanikatis and J. Bailey, 1997, Studies on glycolysis.1.Multiple steady states in bacterial glycolysis, Chemical Engineering Science, 52, 2579-2588

K. Mauch, S. Buziol, J. Schmid, M. Reuss, 2001, Computer Aided Design of Metabolic Networks, Chemical Process Control-6 Conference, Tucson, Arizona

E. Nikolaev, P. Pharkya, C. Maranas, A Armaou, 2005, Optimal selection of enzyme levels using large-scale kinetic models, Proceedings of 16th I.F.A.C. World Congress, Prague, Czech Republic

A. Waechter, L. Biegler, 2006, On the Implementation of an Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, Mathematical Programming, 106, 1, 25-57

A Multi-Resolution Multi-Scale Computational Approach for Characterization and Analysis of Nanostructured Surfaces

Rajib Mukherjee[a], Ahmet Palazoglu[b], Jose A Romagnoli[a]

[a]*Cain Dept. of ChemicalEngineering, Louisiana State University,Baton Rouge 70803, USA*
[b]*Dept. of Chemical Engineering and Materia Science, University of California, Davis, USA*

Abstract

Structures and property of surfaces are very important in different chemical, physical and biological processes. Understanding the surface characteristics in the microscopic level is essential in order to relate the surface characteristics to the performance of the product. Relation of product performance with surface characteristics helps to improve the product performance through optimizing the manufacturing process. Spatial distribution of surface features which defines the surface characteristics can be captured by the multi-resolution capabilities of wavelet transforms (WT) that can provide not only frequency localization but also spatial localization of feature signatures. A multi-scale molecular simulation can help to investigate the physical and chemical mechanism in the surface. Together with the multi-resolution surface feature analysis, the multi-scale molecular simulation will give a better understanding of the surface phenomena and its relation with the performance matrices. In this paper we discuss the application of this approach for surface characterization of Rh(111) in the adsorption desorption of CO. The adsorption on the surface depends on its crystal lattice structures and the presence of defects. In the atomic level a first principle density functional theory (DFT) calculation is used to find the adsorption energy. In the mesoscopic level a kinetic Monte Carlo (KMC) model of the surface is used to simulate the temperature programmed desorption (TPD) from the surface. The on-top adsorption energy increases with surface defects in the form of vacancies which shifts the peak maximum of TPD to a higher temperature. To characterize the surface, fractal dimension of the crystal surface is found using wavelet transformation. The fractal dimension of the surface increases with presence of vacancies.

Keywords : wavelet transform; density functional theory (DFT); kinetic Monte Carlo (KMC); temperature programmed desorption (TPD)

1. Introduction

The performance of various products like heterogeneous catalysts, solar energy conversion panels, semiconductor devices etc. are controlled by the characteristics of the molecular interfaces and nanostructure of their surfaces. The surface structure of these applications, created by molecular deposition on the surface due to surface kinetics and surface diffusion, governs the function of the material. To create an application of desired features, we first have to understand the surface characteristics of these applications in the microscopic level find the relation of product performance with

surface characteristics, thereby improve the product performance through optimizing the manufacturing process towards novel tailored nanostructures. Thus, the key to control fabrication of applications whose performance matrix is governed by its surface feature is to characterize the surface and use those characteristic features as control parameters. In this way we can optimize product performance as well as have a quality control.

The performance of heterogeneous catalysts is controlled by the characteristics of the molecular interfaces and nanostructure of their crystal surfaces. The characterization of these nanostructures often involves surface imaging using Scanning Electron Microscopy (SEM), high resolution electron microscopy (HREM) and different Scanning Probe Microscopy (SPM). A high resolution electron microscopy (HREM) image of the catalyst surface used by Yakaman et al. to obtain the crystal lattice structure of the active metal, revealing the defects on the crystal surface of the crystal lattice structure. Crystal formed in non equilibrium condition will have defects. Such surface defect affects the surface property. The effects of surface defects like steps and kinks in adsorption and subsequent dissociation have been studied for some time In this paper, we have studied the effect of surface vacancies in the adsorption-desorption property. While microscopy images can provide qualitative information about the features of a surface, we have used wavelet transformation on simulated surface with and without vacancies for the quantitative description of the surface morphology. A first principle DFT model of similar surface is used to relate the surface characteristics with its adsorption-desorption property. In our study the adsorption energies of CO in Rh(111) for low surface coverage (p(2×2) for 0.25 monolayer (ML)) were obtained using Car-Parrinello Molecular Dynamics (CPMD) program. These adsorption energies are then used to find the desorption kinetics by the kinetic Monte Carlo (KMC).

The TPD of CO for a regular Rh(111) surface at low surface coverage of CO was compared with the experimental results obtained by Dubios and Somorjai. A surface with vacancies will have higher surface energies. The effects of the presence of vacancies on adsorption energies and the corresponding effects on TPD were simulated. The changes in fractal dimension with the presence of vacancies are noted. A regular surface will have a fractal dimension of two while a surface with vacancies will have fractal dimension between two and three.

2. Computational details

2.1. Surface Fractal dimension

Fractal dimension can be obtained from the variance of wavelet coefficients. The roughness coefficient or the Hurst exponent, H, is obtained from the slope (s) of the plot of wavelet variance at different scale index (Mukherjee et al.). The slope is related to the Hurst exponent as given in Equation 1.

$$H = \frac{s-2}{2}$$

(1)

For a surface the roughness or Hurst exponent is related to the fractal dimension, D, as $D = 3 - H$.

2.2. DFT estimation of adsorption energies

The surface adsorption energy E_{ad} at a specific temperature and pressure is given by the difference in the internal energy of the surface with adsorbate in it $E_{CO/Rh-slab}$ and

sum of the internal energy of the surface $E_{Rh-slab}$ and adsorbate E_{CO} as shown in Equation 2 (Gajdos et al.).

$$E_{ad} = E_{Rh-slab} + E_{CO} - E_{CO/Rh-slab} \qquad (2)$$

A DFT estimation of the adsorption energies of CO in Rh(111) is calculated for both on-top and bridge sites. 0.25 ML surface coverage is used. Lateral interactions can be considered to be small as CO adsorbate-adsorbate lateral interactions at surface coverage lower than 0.33 ML is not significant (Gajdos et al.). The wave function was expanded on a basis set of plane wave with cut-off of 70 Ry. We have used a norm conserving pseudo potential of Trouiller-Martins with Becke-Lee-Yang-Parr (BLYP) functional and non linear core correction (NLCC). A geometry optimization is done. The adsorption is supposed to occur in ultra high vacuum (UHV), hence there is no dissociation of CO molecules upon adsorption as found experimentally and theoretically by Dabois et al.

2.3. Kinetic Monte Carlo (KMC) simulation of the surface

The KMC is a stochastic algorithm that will propagate the system from state to state. The steps used in simulation are constructing the surface and adsorbate, determining the kinetics of desorption process taking the lateral interaction into consideration, followed by the simulation of the elementary processes like diffusion and desorption to determine the output. The intrinsic binding energies, activation energies are derived from DFT results and lateral interaction energies are obtained from the values estimated by Oh et al.. The structure of the catalyst is assumed to be rigid and represented by a lattice with each point corresponding to a surface site. Every lattice site is assigned with the type of the site and whether or not adsorbate CO is associated with it.

2.3.1. Surface Kinetics

We have used the transition state theory to get the kinetic parameters from the activation barriers and the rate constant of the surface features like desorption are described by the Arrhenius equation. The overall rate for a particular reaction r_i is given by Equation 3.

$$r_i = A \exp\left[\frac{-(E_{i-des} + Q)}{k_B T}\right] \qquad (3)$$

The energetic barrier to the desorption process are computed according to the local environment around each site. E_{i-des} is the energy of desorption without lateral interaction and Q is the interaction energy due to lateral interaction. A is the pre-exponential factor and k_B and T are the Boltzmann constant and absolute temperature respectively. The total rate (r_{tot}) of all surface processes is calculated by summing the rates for all elementary reactions ($\sum_i r_i$) in ascending order. A random number (RN) is drawn to determine the event that will occur in the next time step. The forward time step Δt is then calculated by Equation 4.

$$\Delta t = \frac{-\ln(RN)}{r_{tot}} \qquad (4)$$

2.3.2. Surface Diffusion

The diffusion rate for the systems studied was assumed to be fast compared to the other surface process like desorption. Therefore the surface is equilibrated after every time step using a Metropolis algorithm. First an occupied site on the surface is chosen randomly and the binding energy of the initial site Q_i is calculated. Then a vacant location in the vicinity where CO can move is chosen randomly and the CO is moved there. The binding energy of the final site Q_f is calculated. The difference of the two adsorption energy is given by $\Delta Q = Q_f - Q_i$. The probability of acceptance of the move is given by Equation 5.

$$P_{i-f} = \min\left[\exp\left(-\frac{\Delta Q}{kT}\right), 1\right] \tag{5}$$

At the initial surface temperature, the surface is populated by placing a given coverage (p(2×2) for 0.25 ML) of CO. The CO molecules are placed on the on-top sites due to the minimum adsorption energy of the sites. The surface diffusion algorithm is used to equilibrate this initial coverage by performing the diffusion step iteratively. A desorption process is then performed according to its local environment and the time step required in this process is calculated. The process of surface diffusion followed by desorption is continued till the total time at a given temperature exceeds one second. After reaching one second, the surface temperature is increased and the process is repeated. In this way, the numbers of molecules per second desorbing at different temperatures are calculated.

3. Results and discussion

The adsorption of CO on Rh(111) has been studied. The on-top adsorption energy increases in surface with vacancies. The temperature programmed desorption (TPD) of CO from regular Rh(111) surface and that with vacancies is modeled using kinetic Monte Carlo (KMC). The peak maximum of TPD from a regular Rh(111) surface is close to that found experimentally by Dubois et al. for low surface coverage of CO. The shift in the peak maximum of TPD has been observed for a rough surface due to the change in activation energies for desorption with the presence of vacancies.

3.1. Surface Fractal Dimension Estimation

Vacancies are caused by missing atoms which makes the surface rough. A fractal dimension of the surface is created for surface characterization by simulating the surface in a periodic manner. This will keep the fractal dimension same as it is based on the self affine nature of the surface. Simulated surfaces without and with defect are shown Figure 1(a) and 1(b) respectively. The corresponding fractal dimensions is calculated from the slope of Figure 2(a) and 2(b) using Equation 1.

<div align="center">(a)   (b)</div>
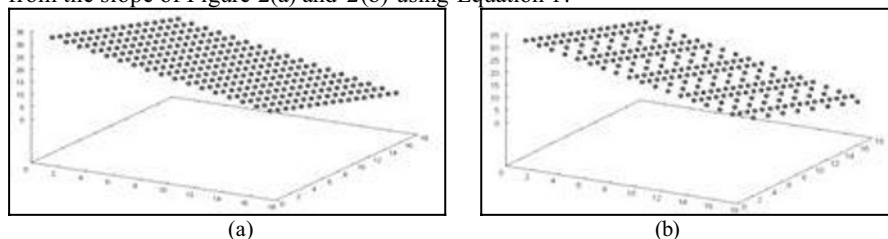
Figure 1 : Simulated (111) lattice plane (a) without/ (b) with defects.

(a)   (b)
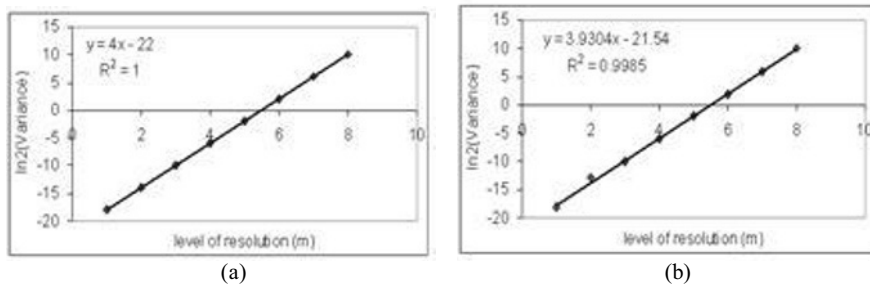
Figure 2 : Wavelet variance vs. different level of resolution for (a) regular (111) lattice plane and (b) plane with vacancies.

The fractal dimension for a regular surface is 2. The fractal dimension for the surface with vacancy is calculated to be 2.035. This shows that the surface energy increased with its fractal dimension.

3.2. Surface Adsorption Energies

The effect of higher surface energies of rough surface is seen in the adsorption energy. A three layer of Rh(111) surface atoms with CO adsorbing on it are modeled. The adsorption energies for on-top sites on regular surface and surface with vacancies are calculated. Figure 3 shows adsorption on on-top sites. For regular Rh(111) the adsorption energies are found to be 32.8 kcal/mole and that on Rh(111) surface with vacancies are 35.1 kcal/mole. This shows higher adsorption energy when compared with surface without roughness. Adsorption energies on bridge sites are also calculated and these adsorption energies are used for TPD.

(a)   (b)
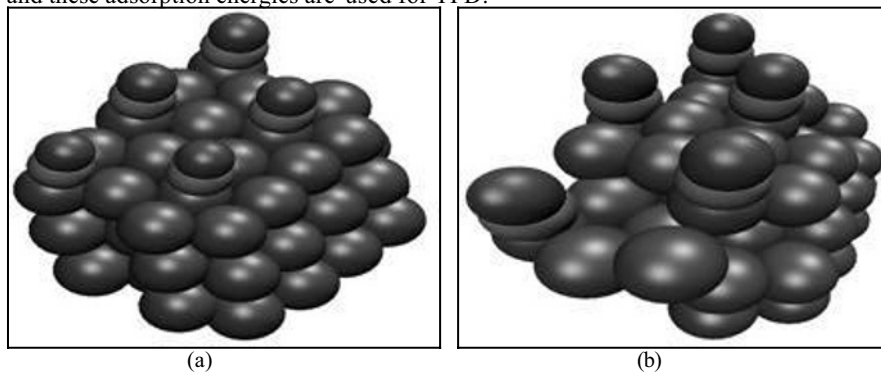
Figure 3: Adsorption on on-top sites in (a) regular Rh(111) and (b) with vacancies.

3.3. Temperature Programmed Desorption (TPD)

The desorption kinetics of CO from the Rh(111) surface is studied under ultra high vacuum (UHV) condition. The low coverage adsorption energies are assumed to be the activation energies for desorption. The CO molecules are arranged in p(2×2) in a regular R(111) surface and the surface temperature is raised at a rate of 25 K/sec. The pre-exponential factor, A, is taken as 5×10^{13} sec^{-1}. The TPD of CO from Rh(111) is shown in Figure 4. The peak maximum is observed at 525 K. This is in agreement with the experimental results obtained by Dubois et al. for low surface coverage indicating that desorption occurs from a single on-top site.
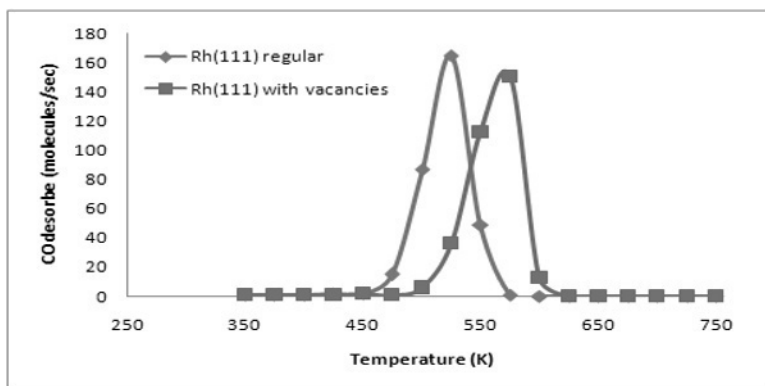
Figure 4: Temperature-programmed Desorption of CO from Rh(111).

The TPD from a surface with vacancies shows a shift in peak maximum from 525K to 575K. The shift in peak maximum to a higher temperature is due to high adsorption energies for a rough surface. We are comparing results from different model surfaces with experimentally created surfaces of similar characteristics.

4. Conclusions

A surface with vacancies has higher energies than regular surfaces. This increases the adsorption activities of the surface. A fractal dimension of the surface estimated in this paper using wavelet transform can be used to characterize the surface in atomic scale. With presence of vacancies, there is an increase in fractal dimension as well as adsorption energies. This increase in adsorption energy shifts the peak maximum from 525 to 575 K in TPD of CO. This shows changes in surface properties with the presence of roughness in form of vacancies. Surface characterization of heterogeneous catalyst in the atomic scale is thus essential for device performance analysis. Similar studies can be made with different lattice surface planes of different active metals as well as defects in terms of dislocations and grain boundaries to see their effects in catalytic activities. Catalytic activities of the surface like dissociation and product formation can also be studied in surfaces with vacancies, dislocations, grain boundaries etc..

References

M.J. Yakaman, G. Diaz and A. Gomez, 1995, Electron Microscopy of Catalysis; the Present, the Future and the Hopes, Catal. Today, 23, pp. 161-199.

CPMD V3.11, copyright INTERNATIONAL BUSINESS MACHINES CORPORATION (1990-2006) and MAX PLANCK INSTTUTE FUER FESTKOERPERFORSCHUNG STUTTGART (1995-2001).

L.H. Dubois and G.A. Somorjai, 1980, The Chemisorption of CO and CO2 on Rh(111) Studied by High Resolution Electron Loss Spectroscopy, Surf. Sci., 91, pp. 514-532.

R. Mukherjee, J.C. Flake, A. Palazoglu and J.A. Romagnoli, 2008, A Multiresolution Spatial Correlation Approach For Line-Edge Roughness Characterization, J. of Wavelet Theory and Applications, Accepted.

M. Gajdos, A. Eichler andJ. Hafner, 2004, CO Adsorption on Close-packed Transition and Noble Metal Surfaces: Trends from ab initio Calculations, J. Phys. Condens. Matter, 16, pp. 1141-1164.

S.H. Oh, G.B. Fisher, J.E. Carpenter and D.W. Goodman, 1986, Comparative Kinetic Studies of CO-O2 and CO-NO Reactions over Single Crystal and Supported Rhodium Catalysts, J. Catal., 100, pp. 360-376.

Optimization of Compressor Networks in LNG Operations

M. M. Faruque Hasan, Md. Shamsuzzaman Razib, I. A. Karimi

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576

Abstract

Liquefied natural gas (LNG) is the most economic way of transporting natural gas (NG) over long distances. Although LNG is an attractive source of clean fossil fuel, it involves energy intensive liquefaction of NG using refrigeration. Often the compressors that run the refrigerant cycles in an LNG plant operate in suboptimal fashion, which results in higher fuel and energy consumption. To this end, we present a generalized model for the compressor operations in multiple interacting refrigerant cycles in LNG and other cryogenic applications. We determine the optimal load distribution between the cycles to minimize total power consumption of the system for a given plant capacity and operating conditions. We also show the applicability of our model using a case study on the AP-X^{TM} LNG process, which includes three interacting cycles, each with single or multiple compressors.

Keywords: Compressor Networks, Refrigeration, LNG Operation, Cryogenics, Optimization, AP-X^{TM} Process.

1. Introduction

Natural gas (NG) is the 'natural' choice among fossil fuels. It is the cleanest with abundant proven reserves. It contains mainly methane (about 90%), ethane, propane, butane, and trace amounts of nitrogen and carbon dioxide. Currently, NG is the world's fastest growing energy commodity and the third largest primary energy source after crude oil and coal. It is also the fastest-growing and second largest energy source for electric power generation. In 2007, NG consumption was 2637.7 million tons oil equivalent, or about 23.8% of the total primary energy consumed worldwide (BP Statistical Review of World Energy, 2008). The consumption is projected to increase by nearly 52% between 2005 and 2030 (International Energy Outlook, 2008).

Most NG reserves are offshore and away from demand sites. An attractive option is to liquefy NG at the source and then transport it as liquefied natural gas (LNG) by specially built ships. When liquefied, the volume of NG reduces by a factor of about 600 at room temperature, which facilitates the bulk transport. In fact, LNG is the most economic way for transporting NG over long distances. Furthermore, more than 90% of the feed heating value in a modern LNG plant is shipped as product LNG (Liu et al., 2001). These along with concerns about energy price and security are making LNG the fuel of the future, with the demand doubling every ten years. In 2007, 226.41 billion

cubic meters of NG was transported as LNG (BP Statistical Review of World Energy, 2008).

However, LNG production is still expensive compared to oil. Production and transportation represent nearly 85% of the cost of delivering LNG to the customer's jetty. A world-scale LNG plant usually consumes about 5.5–6 kWh energy per kmol of LNG produced. Therefore, energy is an immediate concern in LNG industry. NG is cooled to and liquefied at around -163 ºC to produce LNG using energy intensive liquefaction via costly refrigeration. Such refrigeration system involves some of the largest compressors in the world, usually driven by frame-type gas turbines using NG as fuel (fuel from feed or FFF) or electric motors. About 40% of the total operating cost of an LNG plant is due to energy consumption in the compressors that run refrigerant cycles. Therefore, optimizing compressor operation in LNG plants is crucial for reducing energy cost, consumption of FFF, emission of CO_2, etc.

Aspelund et al. (2007) proposed a methodology based on pinch analysis to utilize pressure based exergy (available energy) for sub-ambient processes, such as LNG. Del Nogal et al. (2008) presented an optimization framework for the design of mixed refrigerant (MR) cycles which are suitable for LNG. Shin et al. (2007) proposed a mixed integer linear programming (MILP) formulation for optimizing boil-off gas (BOG) compressor operations in an LNG receiving and re-gasification terminal. They minimized the total average power consumption of the BOG compressors. Selot et al. (2008) presented operational planning model for NG production systems. However, hardly any literature rigorously considers compressor operations for LNG production.

Power consumption in the turbine driven compressors is highly nonlinear and depends on three major factors. First, ambient and cooling water temperatures determine utility consumption, turbine efficiency, and the performance of inter-stage coolers present in the compressor network. Second, the refrigeration loads and hence the power consumption in an LNG plant vary with feed NG composition, flow, temperature, and pressure. Plant operators either run the compressors at their maximum speeds, or adjust the speeds according to the changes in feed and ambient conditions. However, they do it based on their experience and try-and-see approach. This often results in suboptimal power consumption levels. Third, while a pipeline transportation and distribution system for NG has a single network of compressors, modern LNG plants use multiple interacting and complex refrigerant cycles involving highly nonlinear dynamics. For instance, the propane pre-cooled mixed refrigerant (C3MR) process involves two interacting cycles, Casecade Phillips LNG process has three interconnected cycles, Dual MR process has two cycles in series, and Liquefin Axens process uses two interconnected cycles. The complex interaction between the cycles poses a major challenge in the optimization of compressor operations in an LNG plant.

Only recently, Hasan et al. (2009) addressed this problem. They optimized the compressor operations for C3MR process which involves two highly interacting propane refrigerant (PR) and mixed-refrigerant (MR) cycles. They identified propane evaporating pressure in the PR cycle to be the key decision variable that determines the loads of the two cycles. They also obtained the optimum propane evaporating pressure using a nonlinear program (NLP), with the objective to minimize the total power consumption in PR and MR compressors. However, the network complexities and interdependence of the two refrigerant cycles were not addressed rigorously. Moreover, no guideline was provided for the load distribution between the two cycles. Most of all, their model is applicable to C3MR process only, since they did not present any generalized model for any number of refrigerant cycles.

In this article, we present a generalized model for the compressor operations in multiple interacting refrigerant cycles. We determine the optimal load distribution between the cycles to minimize total power consumption in the compressors for a given plant capacity and operating conditions. We also show the applicability of our model using a case study on the AP-X^{TM} process, which includes three interacting cycles, each with single or multiple compressors.

In what follows, we state the compressor network (CN) problem. Next, we present a mixed integer nonlinear program (MINLP) to optimize the CN operations in LNG industry. Finally, we demonstrate the benefits of such an optimized compressor network versus a base configuration for the AP-X^{TM} process.

2. Problem Statement

A plant has N interacting refrigerant cycles ($n = 1, \ldots, N$). Each cycle uses network that involves turbines (gas or steam), single-shaft turbo-compressors, valves, exchangers, etc. and steam at different pressure levels. Let i and j denote the compressors ($i = 1, \ldots, I_n$) and components ($j = 1, 2, \ldots, J_n$) in the refrigerants respectively. Let P, T, and F be the pressure, temperature, and flow of NG respectively and T_f be the LNG temperature. Therefore, the total change of enthalpy of NG is constant. In cycle n, the refrigerant is compressed using compressor C_{in} with inter-stage/after coolers using air/cooling/sea water. The final discharge pressure corresponds to the vapor pressure of the refrigerant at its evaporating temperature, which depends on the air/cooling/sea water temperature (T_{sw}) and hence is fixed a priori. The refrigerant then undergoes a pressure reduction through the JT valves to reach the evaporating pressure at the corresponding evaporating temperature. It is then used to precool/liquefy/sub-cool NG and/or refrigerant(s) by evaporating itself in exchangers (precoolers, MCHE, etc.), which is then sent to the compressors again. Assuming no pressure drop in the exchangers, the evaporating pressure is also the suction pressure to the first operating compressor in the network. With this, CN operation problem can be stated as follows.

Given:
1. NG flow rate, pressure, in/out temperature, and composition
2. Air or sea/cooling water temperature
3. J refrigeration cycles with I compressors and C components

Obtain:
1. Optimal load distribution between the refrigeration cycles
2. Flow of refrigerants
3. In/out pressures of each compressor
4. NG temperature after each refrigeration

Aiming to:
Minimize the total cost of power in the compressors.

Assuming:
1. No pressure drops in coolers
2. Identical compressors in each cycle
3. Constant turbine efficiency

We now discuss our MINLP formulation for the above CN operation problem, and apply it to the case of AP-X^{TM} process with a base configuration. Unless stated otherwise, all indices such as i, j, c, etc. assume the full ranges of their valid values in all the constraints.

3. MINLP Formulation

The model involves two sets of primary decisions. The first selects the compressor that is operating using the following binary variable.

$$x_{in} = \begin{cases} 1 & \text{if compressor } i \text{ in cycle } n \text{ is operating} \\ 0 & \text{otherwise} \end{cases}$$

The more is the load in one cycle, the less are the loads in other cycles, and vice versa. However, if the evaporating pressure of the refrigerant is low, then the load in the cycle and the duty in the exchanger are more and the compression ratio is high. This results in lower loads and hence lower duties to the exchangers in other cycles which would then operate with higher evaporating pressure. This results in less compression ratio for those cycles and hence, less power consumption. Therefore, the second set of decisions is to select the evaporating pressure of the refrigerant or the inlet pressure to each compressor network. Let $P_{1,n}$ and $P_{2,n}$ be the in and outlet pressure of the compressor network in cycle n. While $P_{1,n}$ is the decision variable, $P_{2,n}$ is known and equal to the vapor pressure of the refrigerant at its evaporating temperature in cycle n.

Furthermore, since the flow and pressure of the refrigerant affects the efficiency and power consumption of the compressors, the compression ratio (or, the suction/inlet pressure to each compressor) is equally important. Therefore, we also need to select the in and outlet pressures for each compressor. Let p_{in} ($p_{(i+1)n}$) be the in (outlet) pressure of compressor i in cycle n. Since the surface areas and capacities of the exchangers and compressors respectively are fixed, they might operate only in a range of pressures. Therefore,

$$p_{in}^L \le p_{in} \le p_{in}^U \qquad (1)$$

Depending on p_{in}, all I_n compressors in cycle n may or may not operate. However, if compressor C_{in} operates, then $C_{(i+1)n}$ must operate to increase the pressure continuously, until the final discharge pressure is achieved. Moreover, whenever a compressor operates, it increases the pressure up to its maximum capacity. In other words, if compressor C_{in} operates, then $p_{(i+1)n} = p_{in}^U$. Therefore, to model the compressor selection, we use,

$$P_{1,n} \ge p_{in}^U (1 - x_{in}) \qquad (2)$$

$$p_{(i+1)n}^U x_{(i+1)n} \ge p_{in}^U x_{in} \qquad (3)$$

$$P_{1,n} \ge p_{in} - P_{2,n}(1 - x_{in} + x_{(i-1)n}) \qquad (4)$$

$$P_{1,n} \le p_{in}^U + P_{2,n}(1 - x_{in} + x_{(i-1)n}) \qquad (5)$$

$$p_{in} \le p_{(i-1)n}^U x_{(i-1)n} + P_{2,n}(1 - x_{(i-1)n}) \qquad (6)$$

Let W_{in} be the power consumed in C_{in}. We use,

$$W_{in} \ge \omega_n T_{in}^m M_n \left(\frac{p_i^U}{p_i} - 1 \right) \qquad (7)$$

where, ω_n is a known factor that depends on the compressor efficiency and polytropic constant, T_{in}^m is the known mean operating temperature of compressor C_{in}, and M_n is the variable flow of refrigerant in cycle n.

Let TF_n and TR_n be the temperatures after the exchanger where NG and other refrigerants respectively are precooled / liquefied / subcooled in cycle n. An energy balance around the exchangers must be applied to determine TF_n, TR_n, and M_n. We use Antoine equation to express the vapour pressure and temperature relations. The energy balances and the expressions for pressure, temperature, and enthalpy relations for each cycle are similar to those used by Hasan et al. (2009).

Finally, the objective is to minimize the total power consumption in the system and is given by

$$\min \sum_i \sum_n W_{in} \qquad (8)$$

4. Case Study

We take the AP-X^{TM} process (Figure 1) for LNG as the case study. It has three refrigeration cycles, namely PR, MR, and N2. In the PR cycle, propane is used for pre-cooling NG as well as MR. Propane is compressed using multistage propane compressors and then cooled using sea/cooling water. The final discharge pressure ($P_{2,PR}$) corresponds to the vapor pressure of propane at its evaporating temperature, which depends on T_{sw}. Propane is then flashed to a reduced pressure $P_{1,PR}$ at the corresponding vaporizing temperature. This propane is used to precool NG and MR by evaporating itself, which is then sent to the PR compressors again. MR is further cooled in the MCHE.

After performing a pressure reduction through the JT valves, MR is sent to the shell side of the MCHE from the top and used as the two-phase evaporating refrigerant to liquefy NG. The MR pressure $P_{1,MR}$ can be varied depending on the heat load in the MCHE. Once MR leaves the MCHE, it is compressed in the MR compressors and cooled using sea water. Unlike propane, MR is not totally condensed after passing through sea water aftercooler. However, similar to the PR cycle, sea water temperature influences the MR cycle performance.

In C3MR process, final subcooling of NG is performed in the MCHE and the temperature exiting the exchanger is about -150^{0}C to -162^{0}C. However, in AP-X^{TM} process, the temperature exiting the exchanger is about -115^{0}C and final subcooling is carried out in N2 cycle. In N2 cycle, nitrogen is compressed to a high pressure using nitrogen compressors and then cooled to ambient temperature using sea water. The high pressure nitrogen is then cooled in a N2-N2 heat exchanger by exchanging heat with low pressure nitrogen returning to the compressor after executing cooling duty with LNG. After cooling the high pressure nitrogen is passed through an expander where it is expanded to a lower pressure as well as cooled further to a lower temperature. Then the cooled low pressure N2 executes final subcooling of LNG. The use of nitrogen to subcool LNG offers reduction of total refrigeration load on PR and MR cycles.

We apply our model to a base configuration of AP-X^{TM} process with NG feed at 303 K and 64 bar, desired LNG temperature of 123 K. We use sea water at 303 K as the cooling medium in the interstage- and after-coolers. The compressor networks are consist of 2, 3, 1 compressors for PR, MR and N2 cycles respectively. The computing platform is an AMD Athlon™ 64×2 Dual Core Processor 6000+ 3.00 GHz, 3.00 GB of RAM using CPLEX v.10.0.1 (LP solver), CONOPT v. 3 (NLP solver) and DICOPT

(MINLP solver) in GAMS 22.2. Results show the final load distribution between the PR, MR, and N2 compressors are 22.2%, 56%, and 21.8% respectively. All 6 compressors need to operate at these loads. However, such an optimized operation can achieve more than 7% decrease in total power cost compared to that of the base configuration.

Figure 1. Simplified diagram of AP-X^{TM} process for LNG

5. Conclusion

A generalized model is presented for the optimization of compressor operations in multiple interacting refrigeration cycles which are usually common in LNG, air separation, gas processing, etc. We also determine the load distributions between the refrigeration cycles for minimum power consumption in the compressors. However, the model should be extended to consider variations over time and uncertainties in NG feed flow rate, composition, pressure, and temperature, and seasonal variations in sea/cooling water temperature, etc.

References

A. Aspelund, D. O. Berstad, T. Gundersen, 2007, An Extended Pinch Analysis and Design Procedure Utilizing Pressure based Exergy for Subambient Cooling, Applied Thermal Engineering, 27, 2633-2649

F. Del Nogal, J. K. Kim, S. Perry, R. Smith, 2008, Optimal design of mixed refrigerant cycles, Industrial & Engineering Chemistry Research, 47(22), 8724-8740

M. M. F. Hasan, I. A. Karimi, H. E. Alfadala, 2009, Optimizing compressor operations in an LNG plant, Proceedings of the 1st Annual Gas Processing Symposium, H. Alfadala, G. V. Reklaitis amd M. M. El-Halwagi (Editors), 179-184

Y. Liu, C. E. Lucas, R. W. Bower, J. C. Bronfenbrenner, 2001, Reducing LNG Costs by Better Capital Utilization, Presented in the 13th International Conference and Exhibition on Liquefied Natural Gas (LNG13), Seoul, Korea, 14-17 May, 2001.

A. Selot, L. K. Kuok, M. Robinson, T. L. Mason, P. I. Barton, 2008, A short-term operational planning model for natural gas production systems, AIChE Journal, 54, 495-515

M. W. Shin, D. Shin, S. H. Choi, E. S. Yoon, C. Han, Optimization of the Operation of Boil-Off Gas Compressors at a Liquefied Natural Gas Gasification Plant. Industrial & Engineering Chemistry Research, 46, 6540-6545

1773

Development of a Micro Heat Exchanger Made with Ceramic Multi-Layers (LTCC) and its Setup to Gas Flow Measurements

Elsa Vásquez-Alvarez[a], Francisco T. Degasperi[b], Mario R. Gongora-Rubio[c,d] and Reinaldo Giudici[a]

[1]*Department of Chemical Engineering, Polytechnic School - University of São Paulo - Av. Prof. Luciano Gualberto, n. 380, trav. 3, CEP: 05508-900 – São Paulo – SP – Brazil. (rgiudici@usp.br).*
[2]*Faculdade de Tecnologia de São Paulo – FATEC-SP – CEETEPS – UNESP – São Paulo, SP – Brazil.*
[3]*Instituto de Pesquisas Tecnológicas – CTPP - CEP: 05508-901 São Paulo - SP – Brazil.*
[4]*Laboratório de Sistemas Integráveis, Escola Politécnica da Universidade de São Paulo – CEP: 05508-900 – São Paulo – SP – Brazil.*

Abstract

A green ceramic tape micro heat exchanger was developed using LTCC technology. The device was designed by using a CAD software and 2D and 3D simulations using a CFD package (COMSOL Multiphysics) to evaluate the fluid behavior in the micro-channels. The micro heat exchanger is composed of five thermal exchange plates in cross flow arrangement and two connecting plates; heat exchanger dimensions are 26 x 26 x 6 mm^3. Preliminary tests were carried out to characterize the device both in atmospheric pressure and in vacuum. The same techniques used in vacuum technology were applied to check the rotameters and to prevent device leakages. Thermal performance of the micro heat exchanger was experimentally tested.

Keywords: LTCC, micro structured heat exchanger, cross flow, simulation.

1. Introduction

Global competition challenges the chemical industry to search for technology innovations and, for this to happen, relevant organization units are necessary, such as market and sales, research and development, process technology and engineering. In this context of globalization and sustainability, it is evident that process intensification is a step for the future of the area. Process intensification is the development of new equipment and/or techniques which allow a large improvement in some of the production stages in the chemical industry, increasing the area/volume relation of the system and decreasing resistances to the energy and mass transportation in order to increase the correspondent transfer rates.

The advance in Micro fabrication technology opened a new field of research in the Chemical Process area. These techniques applied to microelectronics created huge new markets and are now applied to Chemical Engineering, creating new horizons for extraordinary developments. Low Temperature Co-fired Ceramics (LTCC) technology

enables the possibility of fabricating 3D devices like holes, channels and hollows in a scale from one micron to few millimeters using multiple-layer green ceramic.

The objective of the current work is the development, characterization and performance tests of a green ceramic micro heat exchanger. The device was designed by using a CAD software and simulated in 2D and 3D using a CFD package (COMSOL Multiphysics) to evaluate the fluid behavior in the micro-channels. The construction was made using the LTCC technology. Several Computer Numeric Control (CNC) machining tests, glass ceramic tapes adherence, device lamination and sintering were made before the fabrication of a 3D device was feasible. The micro heat exchanger is composed of five thermal exchange plates in cross flow and two connecting plates with 26x26x6 mm^3. Firstly, the experimental setup was characterized and performed using tests developed according to the gas flow measurement techniques of vacuum technology, more specifically, conductance. Rotameters were used to measure the N_2 gas flow. The leak tests were performed both in the micro heat exchanger and the gas flow setup junctions. The experimental setup is running and the regime of gas flow is viscous (Knudsen < 0.01).

2. Design and Fabrication

Several different geometry projects were analyzed in order to define the device construction, but only four of them are discussed in this work. The geometries were simulated and the results were compared to decide the best configuration for construction. The micro heat exchanger project presents stacked plates with parallel and rectangular shape channels.

2.1. Micro plate designs

Several variations in the fluid distribution were studied in order to assess and determine the best velocity distribution in each one of the plate channels. Among the various options in terms of channel number, length, width and height, four geometries were projected in the CAD. Each plate has two holes: one for the passage to the next plate and another for the hollow outflow. All evaluated geometries presented parallel micro channels with 15,677 mm length, 350 µm width and 500 µm depth.

The main challenge of the project was the hollow that distributes the fluid to the channels. Hence, Figure 1 shows the fluid flow in the proposed geometries. Thus, the first proposed project (A-type) has a side entrance and a triangular shaped fluid distribution hollow (see Figure 1a). In the second project, the B-type, the fluid inflow happens in the central part of the plate and the distribution is also triangular shaped, according to Figure 1b. In the third proposal (C-type), the fluid distribution is in a circular shape, according to Figure 1c. There is a fourth geometry (D-type) which holds two walls in the edges as in Figure 1d.
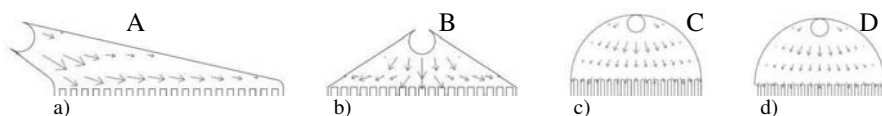
Figure 1. Fluid distribution hollows proposed for the micro heat exchanger plate.

2.2. Fluid behavior in CFD

The geometry project for the device plate involves the fluid dynamics, so simulations of the correspondent processes are necessary to make the device project cheaper, faster, more reproducible and more controllable. In this case, a simplified CFD simulation was

used to assess the effect of the fluid distribution chamber to the channels and determine the best geometry. This simulation was made with the COMSOL Multiphysics v 3.3a, considering nitrogen at 600 K with 1 bar pressure, 0,561 kg/m3 density, 2,95.10-5 pa.s viscosity and 10m/s velocity as the work fluid. To make the simulation easier, it was assumed that the fluid had a top-bottom direction

The assumed boundary conditions were:
- No fluid slipping on the walls
- Zero relative pressure in the outflow and
- Inflow velocities specification,

The incompressibility, the laminar regime and a Knundsen value below 0.01 were adopted as simplifications in the problem, in order for the environment to be treated as continuous (Commenge, 2002). Figure 2 shows the qualitative result of the CFD simulation both in 2D and 3D for the D-type geometry.

a) 2D Simulation   b) 3D Simulation
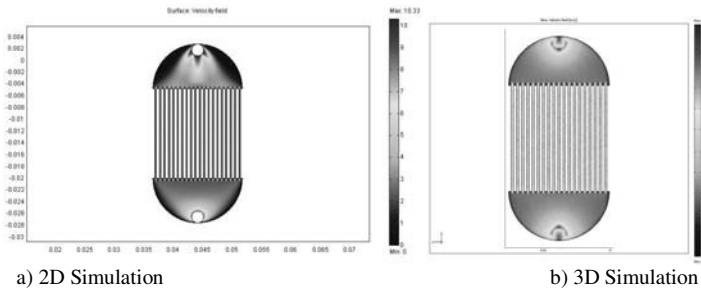
Figure 2. COMSOL Simulations for D-type geometry.

2.3. LTCC Technology

The aluminum vitroceramic in the "green" state (before sinterization) enables the generation of holes, channels and hollows in a simple way up to 200 micrometers. After the individual layer process, they are laminated and a multi-layer system is obtained. When the laminated is co-sinterized, a hard body is generated with the desired application. This technology was initially developed for high reliability, power and frequency applications and it has been an excellent alternative for the silicon in the construction of equipment used in chemical processes, because it allows the generation of 3-D structures with the added possibility of integration of the sensors and thermal actuators. The fabrication process using the LTCC hybrid technology was described in general terms by Gongora-Rubio et al. (2001). Recent researches have shown the integration of microfluidics and microeletronics using the LTCC hybrid technology, such as Martinez-Cisneros et al. (2007) and Ibáñez-García et al. (2008), but a new approach is presented in this work: the development of a micro heat exchanger with multi-plates.

3. Experimental Procedure

3.1. Micro heat exchanger construction

For the fabrication of the micro heat exchanger plates, green ceramic 951 PX tapes with 250 μm thickness from Dupont were used. Since this technology is based on the multilayer approach, the current project considers four stacked layers for a specific geometry. The device construction process follows the sequence presented in Figure 3.

The geometries are projected in the CAD software and, then, they are simulated in a CFD software in order to choose the best fluid dynamics behavior.

In order to stack the tapes, it was necessary to carry out different compaction tests with variable temperatures, time and pressure parameters to prevent delamination, because the channels depth is 500 μm (the same thickness of two tapes). The lamination process was made in a uniaxial hydraulic press with the adaptation of an electric resistance set for the thermal transfer. The microchannel machining was made with a CNC PROTOMAT C100/HF milling cutter, which uses 400 μm cutters and 2.00 mm drills. Then, the green ceramic plates are glued together in a crossed way with a thin layer of an organic fluid (Da Rocha et al. 2004) and they are laminated at a 26 MPa pressure. The micro heat exchanger has three plates for one flow and two plates for the other.
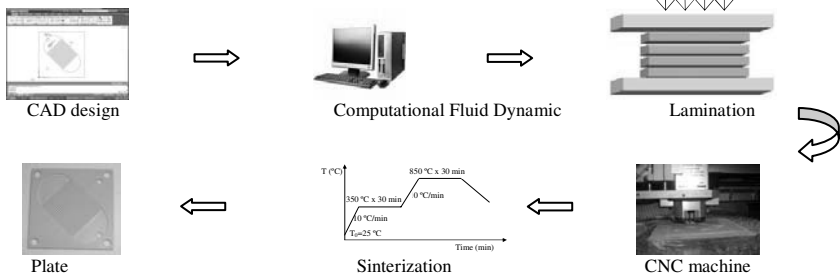
Figure 3. Fabrication process of the micro heat exchanger plates.

3.2. Experimental Setup with Procedures

The cross flow micro heat exchanger (See figure 4a) was previously installed in a metal housing which was specially projected for this experiment, but this material led to huge heat losses. Then, in order to avoid this problem, an acrylic housing was later projected and constructed (see Figure 4b). The micro heat exchanger was attached to the acrylic housing with Kalrez o-rings, which can also be used to avoid heat losses.

a)   b)   c)
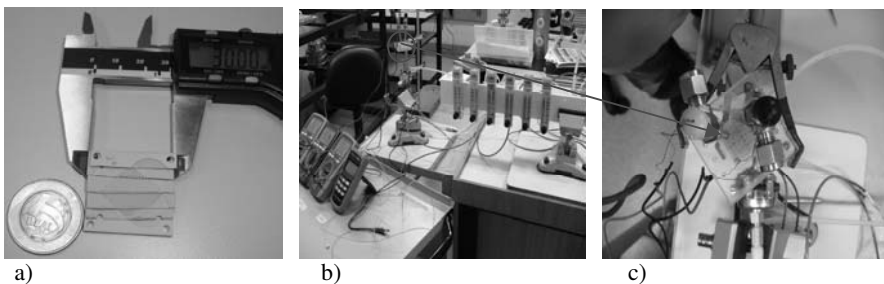
Figure 4. Plates in LTCC and experimental setup of the micro heat exchanger.

The gas flow circuits were considered using N_2 gas, rotameters to measure gas flow, thermocouples to measure the temperatures in each point of gas inlet and outlet in the micro heat exchanger. It was necessary to build a special system of temperature measurement adapted from commercial thermocouples, since the dimensions are extremely reduced. Electrical heater is used to heat up the warm gas stream entering in the exchanger (see Figure 4c) to the desired temperature.

4. Result and discussion

4.1. CFD Simulation

The analysis of the simulations in function of the average and maximum velocities developed in each channel is shown in Figures 5a and 5b, respectively. It indicates that the D-type geometry presents the best flow behavior with a standard deviation in the maximum velocities of 0.0086 m/s and a Re = 54.

a)   b)
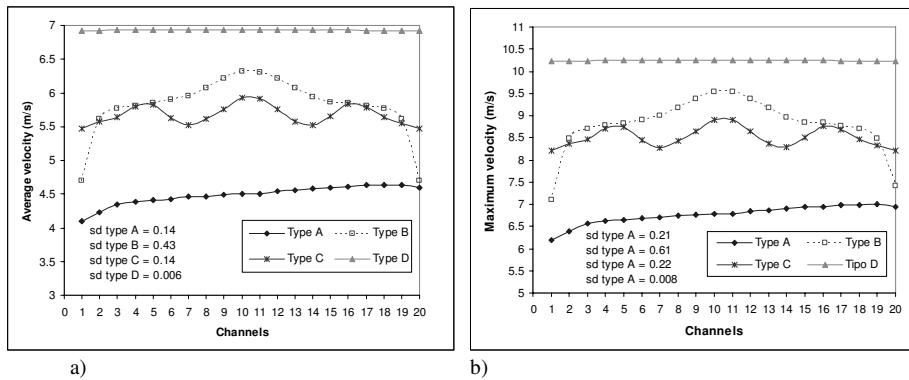
Figure 5. Comparison of the simulation results in the four proposed geometries.

4.2. Experimental results

The preliminary tests were made in atmospheric pressure for different flows and configurations in the micro heat exchanger.

The first experimental test was performed for a flow of 40 cm^3/min of each one of the gases, using the device with the following configuration: three plates of warm fluid and two plates for the cold fluid, the system was tested in co current flow. The temperature measures were made with thermocouples installed in the two inlet flows and in the two exiting flows. It is important to emphasize that the warm fluid is pre-heated before entering in the heat exchanger, by using an independent electrically heated device. The experience was executed for a period of 42 min and the result indicates that the steady state was not attained (Figure not shown). The global energy balances were used to both warm and cold fluid as well as the material properties to calculate the heat changes and confirmed that the wall is in transient regime, with a rate of average temperature of 0.0834 ºC/min.

The second preliminary test presented in this work was performed for the N_2 with 10 cm^3/min in both entering gas streams, with the following configuration: two plates for the warm gas and three plates for the cold gas. The system was tested for the input flow in parallel. The temperatures were measured in four points for 42 min, which shows that the steady state was attained, as can be seen in Figure 6a. By checking the experimental result through global energy balance one can observe that there is a heat loss of 3.75%, this possibly occurs due to changes in the external surrounding. It can be observed that; in this case; the residence time of gas is larger than that in the configuration of test 1. This explains the steady state attained faster for the same time period evaluated.

The third experimental test made with the micro heat exchanger was the 10 cm^3/min in both feed streams using the same configuration of test 2, but in counter flow. The experiment had a duration time of 28 min, and it was observed that the system reached the steady state as show in Figure 6b. The global energy balance was performed in each

one of the circuits. It was observed a heat loss even lower (1.05%) compared with that calculated for the test 2.

In the three experimental tests performed, the temperature changes in the warm and cold flows were not exactly the same, even though both streams were air with the same flow rate. This is possibly due to the insufficient system insulation.

a)   b)
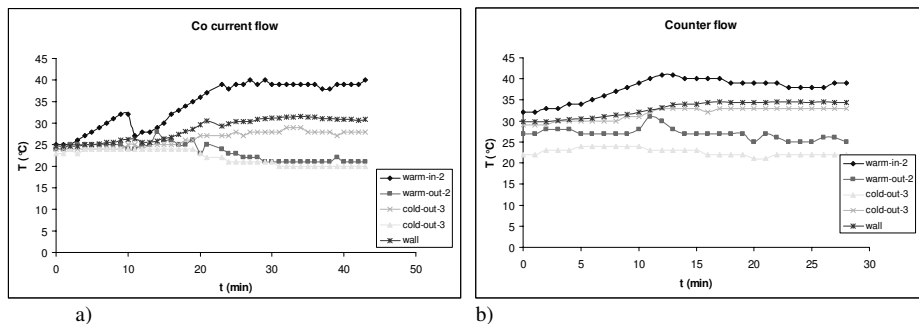
Figure 6. LTCC device experimental results.

Conclusion

LTCC technology is quite useful for developing novel devices to integrate non-electrical functions for microfluidic, analytical chemistry, microplasma and microreaction applications

In this work, a crossflow multi-plate micro heat exchanger was designed, manufactured and experimentally tested. The CFD simulations for one of the plates show the fluid distribution for every channel and the D-type geometry presented the best behavior of flow distribution when compared with the three other geometries considered.

The preliminary tests indicate that the test 3 (counter flow) was the most efficient, and the steady state has achieved faster and with lower heat losses compared with system in parallel flow (test 2).

Further modifications are ongoing in order to improve electronic instrumentation, temperature sensor measurement, flow measurement and system thermal insulation.

Acknowledgements

References

J-M. Commenge, Falk, L.; Corriou, J-P.; Matlosz, M., 2002, Optimal Design for flow uniformity in Microchannel Reactors. AICHE Journal, v. 48, 2, pp. 345 – 358.

Z. M. Da Rocha, Ibañez-Garcia, N.; Oliveira, N. A.; Matos, J. Do R., 2004, Low Temperature and Pressure Lamination of LTCC Tapes for Meso-Systems, In: IMAPS Conference And Exhibition On Ceramic Interconnect Technology, Denver.

M. R. Gongora-Rubio, Espinoza-vallejos, P.; Sola-Laguna, L.; Santiago-Aviles, J. J., 2001, Overview of Low Temperature Co-Fired Ceramics Tape Technology for Meso-System Technology; Sensors and Actuator A-Physical; v. 89, pp. 222 – 241.

N. Ibáñez-García, Martinez-Cisneros, C. S.; Valdés F.; Alonso J. 2008, Green Tape ceramics. New technological approach for integrating electronics and fluids in Microsystems, v. 27, pp. 24-32.

C. S. Martinez-Cisneros, Ibáñez-García N., Valdés F., Alonso J., 2007, LTCC Microflow analyzers with monolithic integration of thermal control, Sensors and Actuator A-Physical, v. 138, pp. 63-70.

Dynamic Simulation of Nuclear Hydrogen Production

Patricio D. Ramírez,[a] Mujid S. Kazimi,[b] Paul I. Barton[a]

[a]*Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 20139, USA.*
[b]*Department of Nuclear Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA.*

Abstract

Hydrogen is an important commodity in the chemical industry, but its production is not CO_2-neutral. Currently, 95% of hydrogen is produced from hydrocarbon feedstocks and it releases 12kg of CO_2 per 1kg of H_2 produced. Using nuclear energy to power water-splitting processes is a cleaner alternative to produce H_2 without emitting CO_2. The operation of such a system is difficult and potentially dangerous, and safety needs to be demonstrated using dynamic simulation. However, current approaches to demonstrate safety cannot deal with the complexity added by chemical reactions. A dynamic process simulator such as JACOBIAN® allows the simulation and study of such a system. This paper describes the implementation of a dynamic model of a nuclear hydrogen production plant in JACOBIAN®, its validation and simulation.

Keywords: hydrogen, nuclear energy, high-temperature electrolysis.

1. Introduction

Hydrogen plays an extensive role in the chemical industry; it is used in the refining of crude oils, the production of methanol and the synthesis of ammonia. Roughly 50 million tonnes of hydrogen are produced worldwide every year, and new applications will require higher levels of production. The conversion of tar sands, shale oil and coal to liquid fuels are examples of growing applications. Hydrogen demand will also increase with the conversion of biomass to liquid fuels, in the search for CO_2-neutral fuels. Other future applications will include fuel cell powered vehicles.

Unfortunately, current hydrogen production is not CO_2-neutral, as it emits large amounts of CO_2 and is based on fossil fuels. 95% of hydrogen is currently produced from hydrocarbon feedstocks, and the production processes release around 12kg of CO_2 per kg of hydrogen. In particular, hydrogen produced in this way cannot be used to create CO_2-neutral fuels derived from biomass. Nuclear hydrogen production can overcome these difficulties. The heat and/or electricity from a nuclear reactor can split water into O_2 and H_2, avoiding the production of CO_2 and the use of fossil fuels. High temperature electrolysis and the Sulphur-Iodine process are considered the most promising technologies to do this, and national laboratories in the U.S., Japan and France are working on their development.

However, the operation of a nuclear reactor coupled to a chemical plant becomes more complex, making it harder to demonstrate the safety of the overall system during transients. Demonstrating safety during transients is needed to gain approval from the regulatory agencies, to get public acceptability, and to achieve good economics and plant functionality. The coupled system must behave safely under different transients:

start-up, shut-down, re-start, off-normal response, and accidents. Normally, the safety of nuclear reactors is demonstrated using simulation codes developed specifically for that purpose, for example, RELAP5 (INL (2006)). In this case though, these simulators cannot handle the complexity added by the chemistry and thermophysical properties of the hydrogen production processes.

This problem was solved using JACOBIAN® (NumericaTech (2005)), state-of-the-art dynamic simulation software, and by developing adequate physico-chemical models. JACOBIAN® is an open modeling environment and simulator, which supports the formulation of models of chemical, physical and biological processes. It allows the efficient simulation of processes, because it solves large-scale, stiff, sparse ODEs and DAEs, provides model and sensitivity analysis tools, and can simulate hybrid discrete/continuous models. A model library was developed to represent a 600MWth nuclear reactor, a heat transfer loop and a high-temperature chemical reaction section for hydrogen production. The models allow the study of different transient scenarios such as start-up and off-normal responses. This paper explains the implementation and results of these studies.

2. Description of Process and Dynamic Process Simulator

The nuclear hydrogen production process considered in this paper uses high-temperature electrolysis to split water. The process contains four components: the nuclear reactor, the gas turbine, the heat transfer loop and the high temperature electrolysis cell (Fig. 1). In this process the nuclear reactor provides the heat for the gas turbine and for the electrolysis cell. The gas turbine produces electricity that will be used to carry out the chemical reactions in the electrolysis cell. Part of the heat produced in the nuclear reactor is transferred to the electrolysis cell by the heat transfer loop. This loop guarantees the safety and operability of the process; it physically separates the nuclear reactor from the hydrogen production plant and decouples their operation by introducing auxiliary heat sources/sinks (not shown in figure). Using heat from the nuclear reactor to perform the electrolysis at high temperatures increases the thermodynamic efficiency of the process.
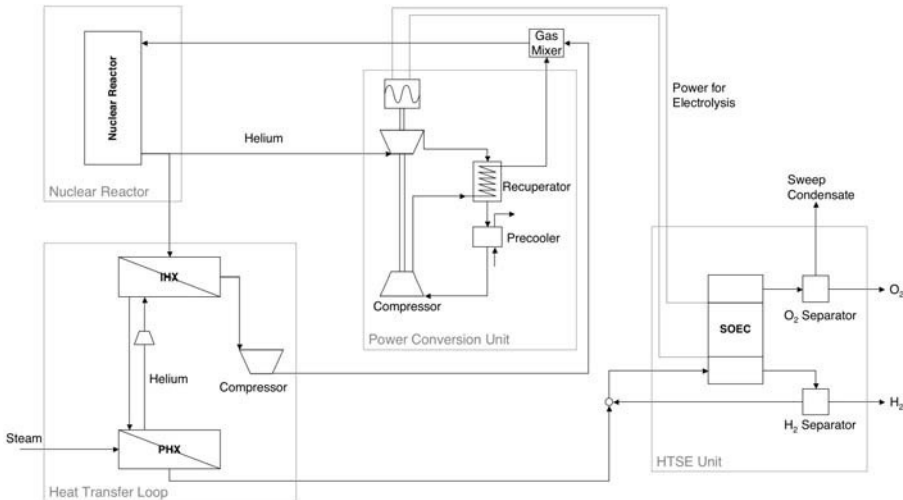
Figure 1. High-temperature steam electrolysis flow diagram.

Equation-oriented process modeling environments can simulate the complexity of such a system in a natural and intuitive way (Pantelides and Barton (1993)). The dynamic behavior of a process can be captured by very detailed models, which can only be constructed in an evolutionary fashion, starting from a simple description and increasing the detail and sophistication when more data and understanding becomes available. Therefore, a dynamic modeling environment, that can offer flexibility in the evolution of models, the possibility of incorporation of discrete and continuous dynamic phenomena simultaneously, and expandability to include other analysis (sensitivity analysis, parameter estimation, optimal control studies) is an enormous help in the design, construction and commissioning of the nuclear hydrogen production process.

JACOBIAN® is an open modelling environment and simulator (NumericaTech (2005)) with these characteristics and it has been selected to simulate nuclear hydrogen production processes. JACOBIAN®, in particular, has state-of-the-art algorithms to identify well-posed systems, to automatically find index, to locate state events, to do hierarchical submodel decomposition, do sensitivity analysis, and to integrate stiff, sparse ODEs and DAEs.

3. Model Formulation

Simplified equations can be used to represent the dynamics of the system, because time scales of less than a second are not important in understanding its operation. This allows to use simpler models with a sufficient level of detail.

In particular, simulating the flow of gas inside the nuclear reactor and inside the heat transfer loop requires a simplified and faster-to-integrate version of the inviscid Navier-Stokes equations. The full system of inviscid Navier-Stokes equations is a hyperbolic system, which means it cannot be integrated efficiently in a dynamic process simulator. However, the fast time scales associated with sound waves can be eliminated from the model, because the focus is the relatively slow transients. This is done by eliminating the transient part in the continuity and momentum equations, and by eliminating the kinetic energy term in the momentum equation, Eq. (1). The resulting system is a parabolic system, which can be integrated in a dynamic process simulator.

$$\frac{\partial \rho v}{\partial x} = 0$$

$$\frac{\partial}{\partial x}(P) = -\frac{2\rho f v |v|}{D} \qquad (1)$$

$$\frac{\partial \rho e}{\partial t} + \frac{\partial}{\partial x}(\rho e v + P v) = -\frac{4U(T - T_{ext})}{D_T}$$

The dynamics of the nuclear reactor are represented using a point kinetics model, a reactivity model and a two-dimensional thermal-hydraulic model (Wang (2003)). These models represent the core neutronics and the heat transfer between the core and the coolant gas. Core neutronics take into account the gross fission power variation due to reactivity change. At the same time, the factors influencing the reactivity are: control rod movement, fuel temperature change and fission product poisoning, as well as external reactivity disturbances. The thermal-hydraulic model represents the heat transfer in the core driven by convection, conduction and radiation. The piping and heat exchangers are also represented by specializing Eq. (1).

The model was built by discretizing the space coordinate. The equations describing the flow of gas in pipes and the heat exchangers were discretized in order to represent the variations along them. 40 nodes were used in the heat exchangers and 20 nodes were used in the pipes. The final model consisted of 1500 variables.

4. Results

The results of the distributed model in JACOBIAN® at the steady-state operating point compare well with the results of the original lumped model design by Davis *et al.* (2005). The values of different variables in the loop showed a discrepancy smaller than 3.5%, and the results of both models are shown in Fig. 2.
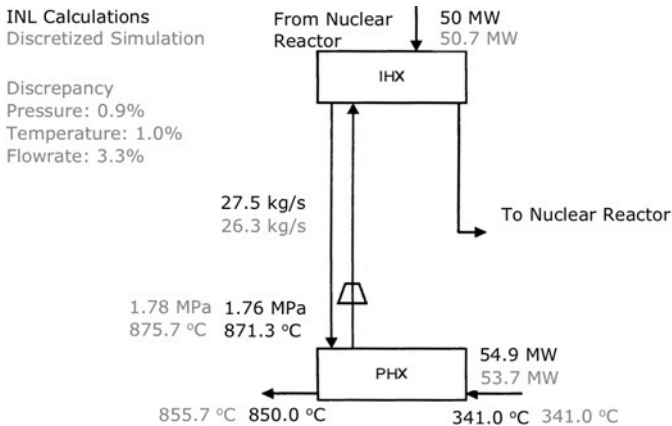
Figure 2. Comparison of results at steady state generated by INL and distributed.

A model in RELAP5 was built to validate the dynamic performance of the distributed model of the heat transfer loop in JACOBIAN®. RELAP5 is a dynamic simulator used to study nuclear reactors, and Fig. 3 shows the structure used to create the model in this platform (based on the work of Davis *et al.* (2005)). At steady state, both models show similar profiles for temperature and pressure.

A simple start-up schedule was simulated as a test for the JACOBIAN® model of the system including the nuclear reactor and the heat transfer loop. The start-up schedule had four stages (Table 1), and the first one was to stabilize the system at a low power level. This was achieved by running the compressors and nuclear reactor at 1% of the designed power level and waiting until the system reached a steady state. Second, the heat removal system was started by increasing the power of the compressors in both loops, and by increasing the flow of gas on the cold side of the process heat exchanger. Third the heat generated in the nuclear reactor was boosted until it reached the design point. Finally, the temperature on the cold side of the process heat exchanger was raised, representing the start-up of the chemical plant.
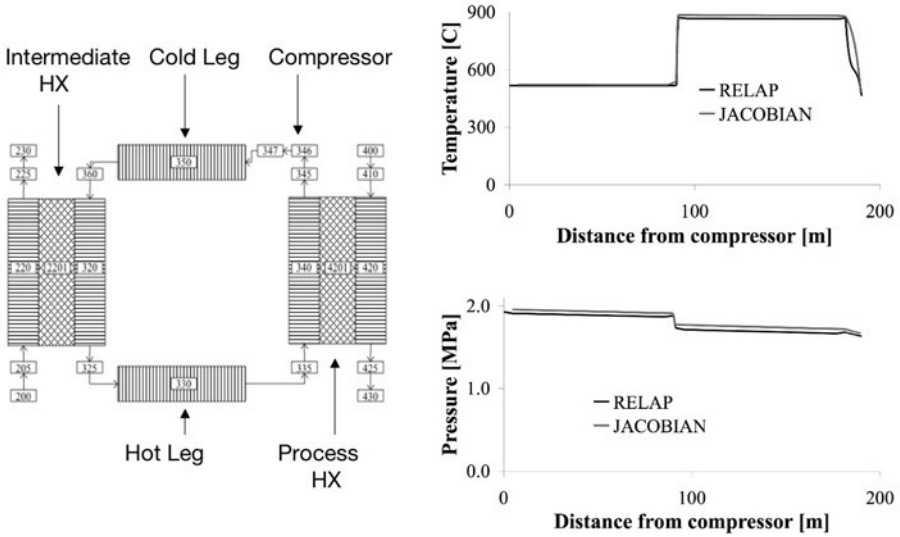
Figure 3. Heat transfer loop for RELAP5 simulation and results at steady state generated by RELAP5 and JACOBIAN® model.

Table 1. Start-up schedule.

Stage	Event	Time
Stabilization	Compressors running at 1%	0 s
Start heat removal	Increase power on compressor in Loop 1	300 s
	Increase pressure on cold side of PHX	
	Increase power on compressor in Loop 2	660 s
Increase heat	Add rod reactivity until QFission = 200 MW	986 s
Connection to chemical plant	Increase temperature in the cold side of PHX	6127 s

Fig. 4 shows the temperature of helium at different points in the system. The first point is at the outlet of the nuclear reactor, the second point is at the cold inlet of the intermediate heat exchanger (IHX). The third point is at the hot inlet of the process heat exchanger (PHX). The helium temperatures at the start of the PHX and at the outlet of the nuclear reactor have a similar profile after 1500s. However, these profiles differ before 1500s; the thermal transients generated by the thermal inertia of the IHX explain this. On the other hand, the helium at the inlet of the IHX is one of the coldest points along the loop. These profiles show the temperature gradients transferring the heat from the nuclear reactor to the chemical plant.

5. Conclusions

Dynamic process simulators, such as JACOBIAN®, can be used to study the operation of chemical systems including a nuclear reactor. In the case of nuclear hydrogen production, the JACOBIAN® model shows very good agreement with the original design from INL and with the model created in RELAP5. The model created was able to simulate the start-up of the system and to show the effect of the thermal inertia of the heat exchangers.
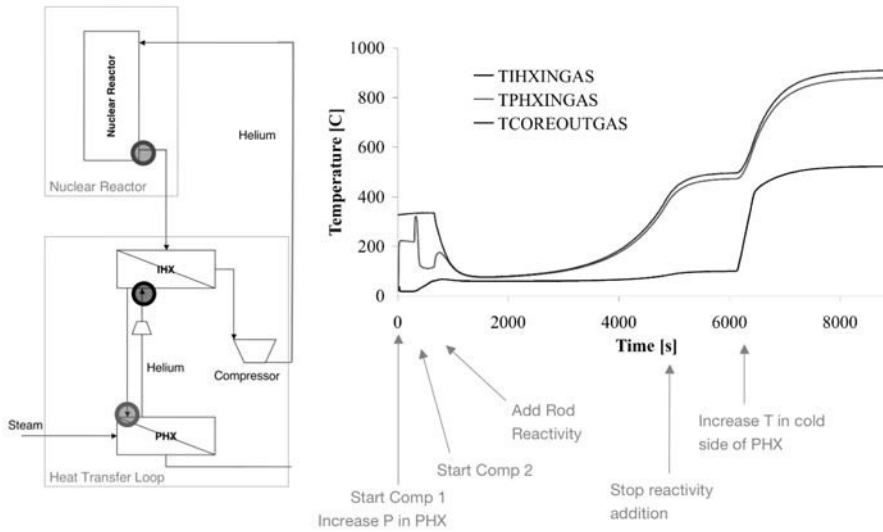
Figure 4. Temperature of helium at various points in the system.

References

C. B. Davis, R. B. Barner, S. R. Sherman, D. F. Wilson, 2005, Thermal-Hydraulic Analyses of Heat Transfer Fluid Requirements and Characteristics for Coupling a Hydrogen Product Plant to a High-Temperature Nuclear Reactor, Technical Report INL/EXT-05-00453, Idaho National Laboratory.

Idaho National Laboratory, 2006, RELAP5-3D Code Manual, INEEL-EXT-98-00834, Rev. 2.4.

C. C. Pantelides, P. I. Barton, 1993, Equation-oriented dynamic simulation current status and future perspectives, Computers and Chemical Engineering, 17, S263-S285.

Numerica Technology LLC, 2005, JACOBIAN dynamic modeling and optimization software, http://www.numericatech.com.

C. Wang, 2003, Design, Anlysis and Optimization of the Power Conversion System for the Modular Pebble Bed Reactor System, PhD Thesis, Massachusetts Institute of Technology.

Acknowledgements

Disclaimer: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Processing of the Atmospheric Distillation Residue with Supercritical CO_2: Conceptual Project

Ana Mehl[a], Raquel S. Macedo[a], Fernando L. P. Pessoa[a], Silvia M. C. da Silva[a]

[a]*Federal University of Rio de Janeiro – Chemistry School - Av. Athos da Silveira Ramos, 149 - Centro de Tecnologia, Bloco E , Sala 201, Rio de Janeiro/RJ, Brasil CEP: 21941-909*

Abstract

Supercritical fluid extraction is a promising technology for treating heavy feed. The present work proposes an alternative route and uses the processes simulator UNISIN/HONEYWELL to evaluate the supercritical extracting process of the atmospheric distillation residue using carbon dioxide as a solvent. The extraction efficiency is a function of pressure, temperature, oil composition, the ratio between solvent and oil rates and can be improved by cossolvent action. This work presents the results of computer simulation studies in the investigation of the capabilities of supercritical carbon dioxide in the extraction of a light fraction from the atmospheric residua (RAT) for the following range of process variables: 50 - 250 bar (operational pressure), $50 – 150^{o}C$ (solvent inlet temperature), a solvent to oil ratio varying from 5:1 to 15:1 and 0-5 mol% of propane as cossolvent. The experimental distillation curve of the feed was used to generate a mixture of pseudo-components representing the RAT. It was observed that the extraction efficiency increases with the solvent to oil ratio and the pressure, whereas it decreases with the increasing of temperature. The RAT vacuum simulation was simulated and the results of both cases were compared.

Keywords: supercritical fluid extraction, CO_2, simulation, heavy oil.

1. Introduction

The current scenario of the global oil production is characterized by the new discoveries and exploration of ultra-heavy and bituminous crude oils. As a result, the refineries should process more and more feeds that exhibit a peculiar behavior in the refining process as well as in the pipeline transportation. Thus, the investigation of technologies to use these fractions, present in greater quantity in the Brazilian oil, is desired. So the development of technology for processing the residue of atmospheric distillation able to withdraw a greater range of products with high added value generating less waste and ensuring that this waste has an appropriate destination is of great importance.

The supercritical fluid extraction may be a viable alternative for the conventional separation process of heavy hydrocarbons in the petroleum industry. The performance of the supercritical fluid extraction in the fractionation of multi-component mixtures depends on several factors among them the solvent density, differences in volatility and intermolecular forces between the solute and supercritical solvent (Hwang et al. (1996)). By use the supercritical fluid extraction technique to treat heavy feeds, products of lower viscosity can be obtained. It represents a cleaner technology wherein it is not necessary to use toxic solvent and the final cuts containing most of the contaminants and asphaltenes can be conveniently discarded (Zhao (2005)).

The atmospheric distillation residuum (RAT) is a fraction with boiling point above $420^{\circ}C$, which is obtained as bottom stream in the atmospheric distillation of the crude oil. It is a residua of high molecular weight, high viscosity, high pour point and low commercial value but it still contains fractions of commercial interest. In the conventional route, RAT is sent to the vacuum distillation followed by the propane desasphaltation in a liquid-liquid extraction. The present work proposes an alternative route and uses the processes simulator UNISIN/HONEYWELL to evaluate the supercritical extracting process of the atmospheric distillation residue using carbon dioxide as a solvent.

The aim of this work is to evaluate the influence of the process variables on the efficiency, using carbon dioxide as the solvent. The effects of the ratio between solvent to oil rate, operational pressure, solvent temperature and the use of propane as an entrainer are investigated.

2. Methodology

The simulation of the RAT extraction with supercritical carbon dioxide was performed using the UNISYM simulator. For comparison purposes the vacuum distillation, which correspond to the conventional refine route, of the same feed was also simulated.

2.1. RAT Characterization

The atmospheric residuum used as the feed of the extraction was analyzed in laboratory and its distillation curve determined by ASTM D2287 is presented in Table 1.

Table 1 – RAT Distillation Curve

Distillation Curve (ASTM D2287) – 760 mmHg, %ml	
IBP	382.2
5/10	410.8/427.2
15/20	441.8/457.5
25/30	475.4/493.9
35/40	512.3/530.9
45/50	549.7/568.2
55/60	587.3/607.1
65/70	627.6/649.1
75/80	673.0/700.8
85/90	726.7/-
95/PFE	-/750.0
(Rec. %m/m)	(87.7)

As in the case of the crude oil, the exact chemical composition of the RAT is unknown, therefore the process simulator uses the distillation curve to define a mixture of pseudo-components to represent the RAT in the simulation. For each of the nine pseudo-components the physical and chemical properties were estimated and the results are presented in Table II. The critical temperature and pressure, the acentric factor and the enthalpy were estimated by the Lee-Kesler correlation and the molecular weight by the Twu correlation. To model the phase behavior of this system the Peng-Robinson cubic equation was selected since it has been extensively used for modeling systems composed by hydrocarbons and light gases. A quadratic mixing rule of van der Walls type was used and the binary interaction parameters were estimated by the simulator. The module chosen to represent the supercritical fluid extraction process was a liquid-liquid extraction column allowing the definition of the number of theoretical stages (N),

where the number of inlet and outlet streams may vary from two (top and bottom) to N. It is necessary to define the pressure profile throughout the column, but in this study the pressure was considered to be constant.

Table II –Pseudo-Componentes Properties

Pseudo-Component	Molecular Weight	Molar Fraction in RAT	Normal Boiling Point(°C)	Density (Kg/m3)	Critical Temperature (°C)	Critical Pressure (kPa)
NBP_470	457.59	0.1105	469.8712	913.3668	618.84	949.5759
NBP_504	504.44	0.1683	503.5198	924.002	645.6333	849.045
NBP_542	562.67	0.1326	541.9196	934.1766	675.166	741.4648
NBP_581	620.71	0.1155	580.8361	948.8383	706.7507	659.6874
NBP_619	685.01	0.1041	619.0583	962.8459	737.4972	587.7264
NBP_657	750.63	0.0914	657.471	976.4508	768.1043	522.5124
NBP_695	815.73	0.0783	695.4081	989.4991	798.0915	464.496
NBP_736	884.89	0.0742	735.5568	1002.996	829.6286	409.4713
NBP_778	962.98	0.1252	777.539	1018.055	863.0186	360.584

2.2. Solvent Selection

Due to their reasonable costs and favorable values of the critical and transport properties, CO$_2$, methane, ethane and propane are the most used solvents in the process that use supercritical technology. The selection of CO$_2$ is linked to some of their characteristics (low toxicity, non-flammable, low critical pressure and temperature and low cost), and mainly to the environmental issues, since in this process, CO$_2$ generated in the refinery will find a suitable destination and no longer represent a further source of emissions of greenhouse effect gases. In addition, the supercritical technology fits the label of "clean technology" since it leads to the minimization of energy consumption and allows the use solvents with low toxicity.

Cossolvent use can improve the extraction efficiency. Hydrocarbons with low molecular weight solubilize simultaneously the paraffinic and iso-paraffinic chains and promote the precipitation of resins and asphaltenes. As a low molecular weight hydrocarbon propane presents excellent solvency power added to a good selectivity. For these reason, the use of propane as an entrainer is analysed.

2.3. SFE Fluxogram

In the supercritical extraction a substrate is separated from the mixture through a dense pure gas or a mixture of dense gases employed as extraction agent. Downstream the extractor, the targets component are separated from the supercritical solvent by means of depressurization. This technique represents a promising alternative concerning high product quality and recovery. Moreover this technique is considered a clean technology due to the possibility of the minimization of energy consumption in the solvent recovering that is recycled to the process. The Figure 1 presents the fluxogram of the supercritical CO$_2$ extraction of RAT process. The supercritical solvent feed the SFE column by the bottom. Downstream the extractor, light fraction recovered from RAT is separated from the supercritical CO$_2$ by means of depressurization. The residuum of the process that can be used as asphaltic residuum leaves the extractor by the bottom outlet.
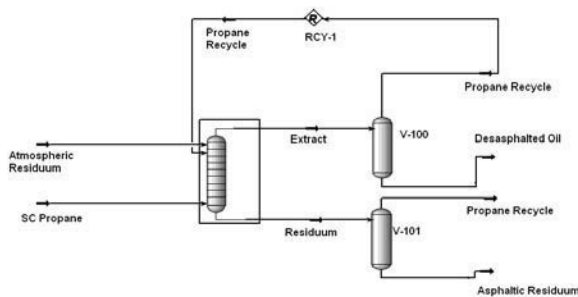
2.1.1. Figure 1 – Process Fluxogram

3. Results and Discussion

To evaluate the extraction efficiency the parameter representing the relative quantity of the recovered fraction of interest within the original feed was chosen as represented by equation (1). Note that the interest fraction is the sum of the three lighter pseudo-components.

$$\%Light\_Fraction\_\text{Re}\,\text{cov}\,ered = \frac{ifm\text{Pr}\,od}{ifmRAT}\,x100 \quad (1)$$

where:

- $ifmProd$ = sum of the mass of the lighter *pseudo-components* (NBP_470, NBP_504 and NBP_542) present in the extract stream.
- $ifmRAT$ = sum of the mass of the 3 lighter *pseudo-components* (NBP_470, NBP_504 and NBP_542) present in the Atmospheric Residuum stream.

The efficiency of the separation is related to the amount of the light fraction recovered from RAT. To evaluate the individual influence of the process variables on its efficiency all variables are kept fixed while one is varied according to a scale.

The effect of the presence of small amount of propane is showed in Figure 2. According to the results the action of propane improves the efficiency of the extraction.
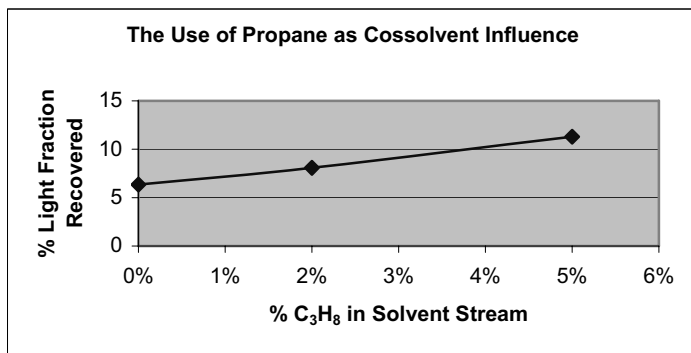
Figure 2 – The Use of Propane as Cossolvent Influence

According to the results presented in given in the Figure 3 the higher the ratio solvent to oil rates the higher the separation efficiency. The composition of solvent is 95 mol% of carbon dioxide and 5 mol% of propane. With respect to the process pressure a direct

dependence was observed since the relative quantity of the light fraction recovered increases with the pressure as shown in Figure 4.
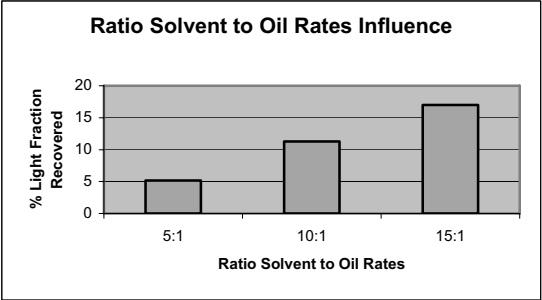
Figure 3 – Ratio Solvent to Oil Rates Influence
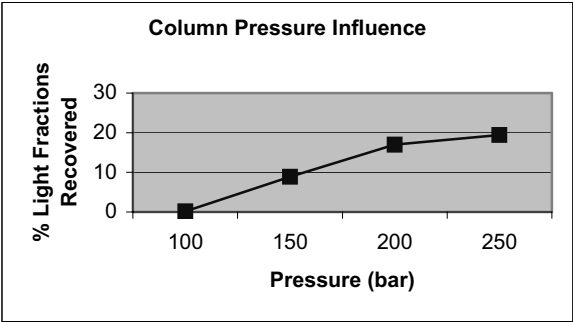
Figure 4 – Column Pressure Influence

With respect to the influence of the solvent feed temperature in the recovery results (Figure 5), the increase in temperature results in the decrease of the relative quantity of the recovered light fraction. This phenomenon can be explained as follows: In an isobaric process the increase of temperature results in the decrease of the solvent density, calculated at 250 bar and 50ºC, 100ºC and 150ºC by quadratic mixing rule of van der Waals type (table V), yielding the reduction of its solvency power and the reduction in the solubility of the solute in the solvent as well.
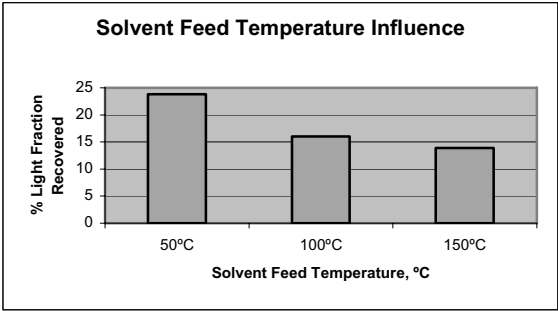
Figure 5 – Solvent Feed Temperature Influence

Table V – Solvent Density

Temperature,ºC	Density, Kg/L
50	0.8134
100	0.5696
150	0.4097

The table VI presents the results of: (i) simulation of the RAT supercritical extraction (solvent composition = 95 mol% CO_2, solvent to oil ratio = 15:1, operational column pressure=250 bar, solvent feed temperature = 50ºC) and (ii) simulation of the RAT vacuum distillation (operational pressure = 0.3 bar, temperature range= 380 to 500ºC). CO_2 in supercritical condition, has the capability to separate the light fraction present in RAT but RAT vacuum distillation showed a best efficiency to recover the light fraction present in RAT.

Table VI – Simulation Results

Stream	RAT	SFC Product	Vacuum Dist. Prod.
NBP-470 Mass Flow (kg/h)	75.2577	33.4999	70.6233
NBP-504 Mass Flow (kg/h)	126.5083	29.3477	36.8513
NBP-542 Mass Flow (kg/h)	111.0476	11.4463	4.8776
NBP-581 Mass Flow (kg/h)	106.7038	4.7932	1.5798
NBP-619 Mass Flow (kg/h)	106.1355	1.9877	0.6015
NBP-657 Mass Flow (kg/h)	102.1138	0.7524	0.2216
NBP-695 Mass Flow (kg/h)	95.0642	0.2673	0.0788
NBP- 736 Mass Flow (kg/h)	97.7244	0.0962	0.0288
NBP-778 Mass Flow (kg/h)	179.4447	0.0595	0.0177
Total Mass Flow (kg/h)	1000	**82.25**	**114.88**

4. Conclusion

In the analysis of the influence of process variables, through simulations has results revealed that: pressure and the ratio between solvent to oil rate presented a direct relation with process efficiency but the solvent feed temperature showed an inverse relation. Despite the fact that all the binary interaction parameters were taken, with the simulator, with the equal value. CO_2, in supercritical condition, has the capability to separate the light fraction present in RAT but the RAT vacuum distillation showed a best efficiency to recover the light fraction present in RAT. It is important to emphasize that further studies should be performed in order to investigate other mixing rules with different binary interaction parameter and solvent mixtures. In this regard, an experimental study in lab scale will be carried out by our group and the results will be compared to those obtained in this work.

References

Hwang J.; Milind D. D.; Hanson F. V.; Dynamic behavior of supercritical fluid extraction of a crude oil and its vacuum residue; Fuel. 1996, 75, 13, 1591-1595.
Zhao S.; Xu Z.; Xu C.; Chung K. H.; Wang R.; Systematic characterization of petroleum residua based on SFEF; Fuel, Vol. 84, nº 6, April 2005, 635-645.

Simultaneous Flowsheet Optimization and Heat Integration of a Bioethanol Processor for PEM Fuel Cell System

Javier A. Francesconi, Diego G. Oliva, Miguel C. Mussati and Pio A. Aguirre[*]

INGAR - Instituto de Desarrollo y diseño (CONICET-UTN), Avellaneda 3657, Santa Fe, CP S3002GJC, Argentina

E-mail addresses: {javierf, doliva, mmussati, paguir}@santafe-conicet.gov.ar

[*]*Corresponding author*

Abstract

The conceptual design of a bioethanol processor for fuel cell applications with heat integration was analyzed via model-based optimization formulation. Basically, the system consists of steam reforming of bioethanol and hydrogen purification process that are coupled to a proton electrolyte membrane fuel cell (PEMFC). The system was implemented within General Algebraic Modeling System (GAMS). The model simultaneously handles the problem of optimal heat integration while performing the optimization of process flowsheet. A 50 kW power generation system is presented as case study; the objective was to obtain the operative conditions for the process units that maximize the system efficiency. The operating variables of the system considered as decision variables were: system pressure, water-ethanol molar ratio and reforming temperature, input temperatures to hydrogen purification reactors, fuel cell temperature, and fuel cell hydrogen and oxygen utilizations. The results obtained predict a global efficiency of 43%, about 5% higher than the reference case. These results demonstrate the importance of simultaneous optimal heat integration for fuel cell-based processes.

Keywords: fuel cell system, bioethanol reforming, heat integration, optimization

1. Introduction

In the last years, fuel cells have received a growing attention for power generation owing to their potential higher thermal efficiency and lower CO_2 emissions per unit of power produced. Many research and development projects have been conducted on both the fuel cell itself and the fuel processors for generating hydrogen. There exist several routes for hydrogen production from the primary fuels. Ethanol presents several advantages related to natural availability, storage and handling safety, ethanol can be produced renewably from several biomass sources. Besides the ethanol-to-hydrogen system has the significant advantage of being nearly CO_2 neutral, since the produced carbon dioxide is consumed for biomass growth, thus offering a nearly closed carbon loop (Benito et al., 2005;Francesconi et al., 2007).

Proton exchange membrane fuel cells (PEMFC) for stationary and mobile applications are highly integrated systems including fuel processing, fuel cell itself and post-combustion units (Godat and Marechal, 2003; Xu et al., 2006). Exhaustive heat integration within the system is necessary to achieve acceptable net electrical efficiency levels. Francesconi et al. (2007) examined the heat integration of the ethanol processor by means of a parametric study using simulation process software. The efficiency

reported was of 38%. The aim of this work was to analyze the simultaneous flowsheet optimization and heat integration of an ethanol fuel processor coupled to a polymeric fuel cell by means of mathematical programming techniques. In this work, the material and energy balances of the system were implemented in General Algebraic Modeling System (GAMS). GAMS is a high-level modeling system for mathematical programming problems. The nonlinear programming model of the system was optimized with CONOPT2 solver. The model simultaneously handles the problem of optimal heat integration while performing the optimization of process flow sheets. The NLP formulation proposed by Duran and Grossmann (1986) is used in order to keep maximum heat integration. This allows analyzing the system energy integration by means of the process integration method (or pinch technology), identifying the most successfully heat exchange opportunities and to define the optimal operating conditions of the ethanol processor with the aim of obtaining the best overall efficiency considering the plant balance.

2. Description of the Ethanol Processor
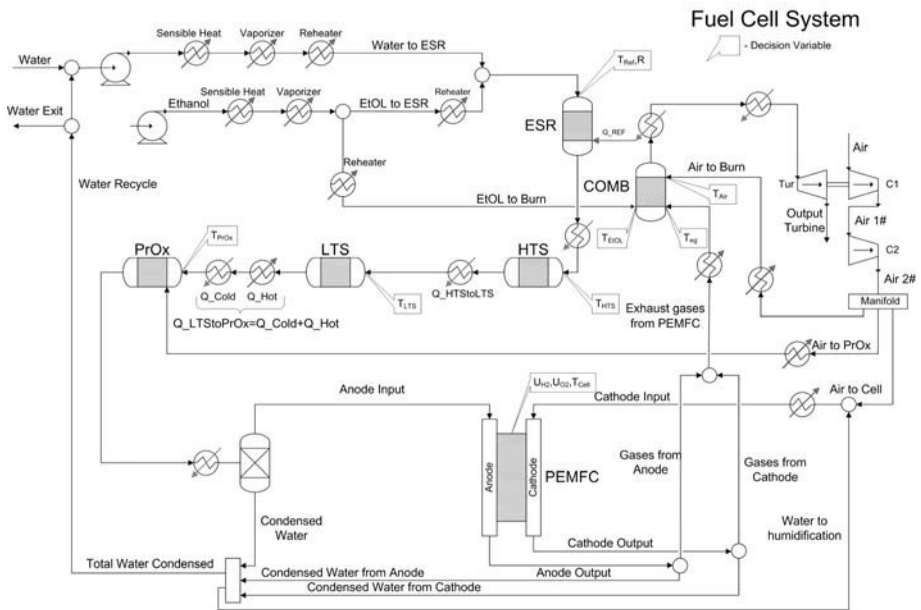
Figure 1. The fuel cell system components and their heat exchange needs.

The Fuel Cell System (FCS) includes a fuel processor which chemically converts ethanol to hydrogen, hydrogen cleanup equipment, a fuel cell stack which electrochemically converts the hydrogen energy to electric power, associated equipment for heat, air and water management, and auxiliary equipment such as pumps and blowers. The fuel processing process has been built according to Francesconi et al. (2007). Figure 1 shows the main components: ethanol steam reformer (ESR), high (HTS) and low temperature (LTS) water gas shift reactors, a preferential oxidation reactor of CO (PrOx), a proton exchange membrane fuel cell (PEMFC), a combustor, compressors and an expander. The major hypothesis regarding the modeling aspect of each unit, and the decision variables considering in the present study are described in the next subsections.

2.1. Ethanol Steam Reforming

This unit was modeled as an equilibrium reactor operating isothermally that involves ethanol steam reforming (feed is composed of fuel and steam). The reformer performs the following reactions according to the mechanism proposed by Benito et al. (2005) for a Co/ZrO_2 catalyst:

Ethanol dehydrogenation: $C_2H_5OH \rightleftarrows CH_3CHO + H_2$ $\Delta H^0_{298} = 71$ kJ/mol (1)

Acetaldehyde decomposition: $CH_3CHO \rightleftarrows CH_4 + CO$ $\Delta H^0_{298} = -21.9$ kJ/mol (2)

Methane steam reforming: $CH_4 + H_2O \rightleftarrows CO + 3\ H_2$ $\Delta H^0_{298} = 206.1$ kJ/mol (3)

Water Gas Shift (WGS): $CO + H_2O \rightleftarrows CO_2 + H_2$ $\Delta H^0_{298} = -41.1$ kJ/mol (4)
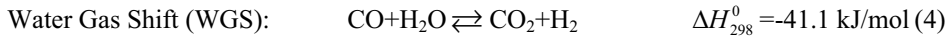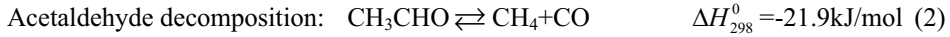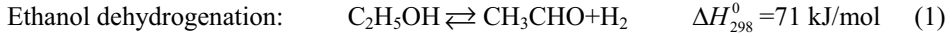
The net energy balance for this complex reaction system is endothermic, and heat has to be supplied from an external energy source in order to maintain the temperature in the reactor. In this study, the reforming temperature (T_{Ref}) and the water to ethanol molar ratio (R) have been considered as decision variables.

2.2. Hydrogen purification section

The CO produced from reforming reactions must be brought down to ppm levels because it gets adsorbed on the noble catalyst of the PEMFC and poisons it. This task is partially accomplished by water gas shift reactors. The water gas shift reaction is carried out in two adiabatic shift reactors in series with an inter-cooler in between to remove the heat of the exothermic reaction. The fist-stage reactor typically operates at 350-500ºC and is called the high-temperature shift (HTS) reactor. The HTS reactor uses a chromiun-promoted iron oxide catalyst. The second stage is a low-temperature shift (LTS) reactor which operates at 150-250ºC, using a copper-zinc catalyst supported on alumina. Performance of these units depends on input concentration of CO and H_2O/CO ratio. Besides, input temperatures affect performance because they modify the output temperature varying the CO conversion.

Chemical equilibrium limits the conversion achieved in the WGS reactor, thus, final CO cleanup occurs in a preferential oxidation (PrOx) unit in which the desired reaction is the oxidation of carbon monoxide (reaction 5). Unfortunately, the selectivity of the catalyst will not avoid the combustion of some hydrogen in the gas stream (reaction 6).

Monoxide oxidation: $CO + O_2 \rightarrow CO_2$ $\Delta H^0_{298} = -283$ kJ/mol (5)

Hydrogen oxidation: $H_2 + \frac{1}{2}O_2 \rightarrow H_2O$ $\Delta H^0_{298} = -242$ kJ/mol (6)
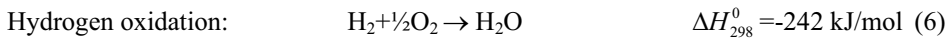
In the model, the CO content at the reactor outlet is fixed at 10ppm. The air flow rate is computed as a function of the CO flow rate assuming two mole of O_2 per mole of CO. Hydrogen combustion is computed considering a complete remotion of the oxygen excess. An adiabatic operation has been considered for the PrOx reactor. The input temperatures of the HTS (T_{HTS}), LTS (T_{LTS}) and PrOx (T_{PrOx}) reactors have been considered as decision variables. Constrains were incorporated according to the operating temperature range of each unit.

2.3. Proton Exchange Membrane Fuel Cell Stack Model

The fuel cell stack was modeled according to Godat and Marechal (2003). The cell voltage (V_{Cell}), cell power (P_{Cell}) and the fuel cell cooling requirement (Q_{Cell}) were computed as follows:

$$V_{Cell} = E^=_{Rev}\left(T_{Cell}\right) + \frac{R_g T_{Cell}}{2F}\left[\ln\left(\frac{P^{in}_{H_2,a} + P^{out}_{H_2,a}}{2}\right) + \frac{1}{2}\ln\left(\frac{P^{in}_{O_2,c} + P^{out}_{O_2,c}}{2}\right)\right] - \eta \tag{7}$$

$$P_{Cell} = I_{Cell} V_{Cell} = 2F \left(f_{H_{2,a}}^{in} - f_{H_{2,a}}^{out} \right) V_{Cell} = 2F U_{H_2} f_{H_{2,a}}^{in} V_{Cell} \tag{8}$$

$$Q_{Cell} = \sum_{i=1}^{inlets} f_i h_i \left(T_{Cell} \right) - \sum_{o=1}^{outlets} f_o h_o \left(T_{Cell} \right) - P_{Cell} \tag{9}$$

The PEMFC was supposed to be isothermal and isobar. Here, E_{Cell}^0 is the reversible voltage adjusted to the cell temperature (T_{Cell}), and the cell voltage was evaluated averaging the partial pressures of the H_2 and O_2 (arithmetic mean) between the inlet and outlet conditions. The operating voltage was chosen as the power level at which unit cell voltage drops $\eta = 0.5V$ from the reversible voltage. The current (I_{Cell}) is related to the hydrogen molar flow rate at the anode (eq. 8). The cell temperature (T_{Cell}), fuel utilization (U_{H_2}) and oxygen utilization (U_{O_2}) were considered as decision variables.

The inlet oxidant to the cathode is humidified to a relative humidity of 80% and the anode inlet stream is always fed at saturated condition. In order to compute the PEMFC fuel cell cooling, an energy balance was made over the cell from the inlet to the exit conditions (eq. 9).

2.4. Post-Combustion System

The depleted fuel of the PEMFC, formed by cathode and anode outlets previous separation of condensate water, is burn in the post combustion system. The generated heat is used to balance the energy requirement of the fuel processing section. Supplementary firing of ethanol is considered if the energy content of the depleted fuel is not sufficient to satisfy the balance. Complete and stoichiometric combustion for the burner unit has been assumed. The preheating temperatures of the ethanol to burn (T_{EtOL}), exhaust gases from the fuel cell (T_{eg}), and the air supply to the combustor (T_{air}) were considered as decision variables.

2.5. Objective function

The objective is to maximize the net efficiency of the system. The FCS global efficiency (φ) is defined as the net energy output of the system obtained from the gross output by subtracting the electrical energy needed to operate FCS auxiliaries such as pumps and compressors divided by the lower heating value (LHV) of the ethanol consumed in the fuel processor for reforming and burning (f_{EtOL}).

$$\varphi = \frac{P_{Cell} - \left(P_{Pumps} + P_{Comp} - P_{Tur} + Q_C/25 + Q_H/0.95 \right)}{\left(f_{EtOL} \times LHV \right)} \tag{10}$$

Q_C is the heat removal from the fuel cell system; it represents the cold utility of the process. We considered a cooling system efficiency of 25 (Larminie and Dicks, 2000). The heat utility (Q_H) is represented by an electrical heater with an efficiency of 0.95.

2.6. Optimization Problem and Heat Exchange Model

The heat exchange network is not considered directly in the model; instead, pinch technology can be used to model the integrated heat exchange system without having to impose a heat exchange network structure. The method is based on Duran and Grossman (1986) approach, it comprises including a set of constraints into the nonlinear process optimization problem so as to insure that the minimum utility target for heat recovery networks is featured. The proposed problem of this optimization with heat integration is given by the following model:

$$\min \varphi = f(x, Q_C, Q_H)$$
$$s.t. \quad h(x) = 0 \tag{11}$$
$$g(x) \le 0$$

$$Q_H + Q_C + \sum_{i=1}^{n_H} Q_i + \sum_{j=1}^{n_C} Q_j = 0 \tag{12}$$

$$\sum_{i=1}^{n_H} F_i Cp_i \left[\max\left\{0, T_i^{in} - T^p\right\} - \max\left\{0, T_i^{out} - T^p\right\} \right]$$
$$\forall p \in \Pi \tag{13}$$
$$-\sum_{j=1}^{n_C} f_j Cp_j \left[\max\left\{0, t_j^{out} - \left(T^p - \Delta T_{\min}\right)\right\} - \max\left\{0, t_j^{in} - \left(T^p - \Delta T_{\min}\right)\right\} \right] \le Q_H$$

where φ is the objective function, Q_H is the requirement of the hot utility, Q_C is the requirement of the cold utility, $h(x)$ is the flow sheet model, $g(x)$ represents the inequality constraints, j is the index for cold streams in the flow sheet, i is the index for hot streams, t_j is the temperature of cold stream j, T_i is the temperature of hot stream i, $F_j Cp_i$ and $f_j Cp_j$ are the heat capacity flow rates of the hot an cold streams, respectively. And the set Π defines the pinch candidate temperatures as follows:

$$T^p = \begin{cases} T_i^{in} & \text{if candidate p is hot stream i} \\ t_j^{in} + \Delta T_{\min} & \text{if candidate p is cold stream j} \end{cases} \tag{14}$$

In this representation the hot duties are considered negatives ($Q_H, Q_i \ge 0$), and cold duties positives ($Q_C, Q_j \ge 0$). In order to avoid difficulties with the NLP solver the smoothing approximation for the max operator cited in Xu et al. (2006), was utilized.

This formulation requires defining, a priori, the cold and hot streams. However, as the flows, compositions and temperatures of the streams are variables it is possible that some stream initially considered hot, could be cold in the optimized case. This behaviour can be handled considering that a specific energy duty is the sum of a cold and a hot stream. In other words, for modeling a given energy duty both possibilities are considered concurrent by means of one cold y and one hot stream, in the optimized case one of them will be null and its energetic requirement will be zero. In the system under study, conditioning the stream incoming at the PrOx reactor can present this situation, and only this stage requires this double representation (see Figure 1). With this approach, was possible to compute the influence of the decision variables so that the best system configuration can be then defined.

3. Results and discussion

The system was optimized considering a net power target of 50kW. It is intended to study the influence of the following operating variables on the efficiency of the integrated system: operating conditions of the reformer (T_{Ref} and R), effect of input temperatures of hydrogen purification reactors (T_{HTS}, T_{LTS} and T_{PrOx}), combustion preheating (T_{EtOL}, T_{Air}, T_{eg}), and operating conditions of the fuel cell stack (T_{Cell}, U_{H2} and U_{O2}). In addition, the pressure of the system (P) was considered as a decision variable, nevertheless pressure drops have been neglected and all streams were defined at the same operating pressure. Table 1 shows the values of the decision variables obtained with its respective upper and lower bounds. The results are compared with the data reported by Francesconi et al. (2007). The global efficiency was increased achieving a 43%. As expected, the hot requirement of the system is satisfied by hot gases post-

combustion, and the hot utility is null. The cool utility achieves the value of 60.5 kJ/h. It is worthy to note, that in the reference case the stream between LTS output and PrOx input represent a cold stream with a duty of 34.4 kJ/h, while at the optimized case this represents a hot stream with a value of -5486 kJ/h.

Table 1. Results obtained from the model-based optimization.

	Reference Case (Francesconi et al., 2007)	present work Value (Lower Bound-Upper Bound)	
Efficiency (%)	38.0	43.0	
P (atm)	3.0	2.2	(1.5-5)
R	4.0	3.0	(2-6)
T_{Ref} (ºC)	709.0	714.3	(550-900)
T_{HTS} (ºC)	350.0	365.8	(350-550)
T_{LTS} (ºC)	150.0	163.6	(150-250)
T_{PrOx} (ºC)	237.0	178.6	(150-300)
T_{Cell} (ºC)	80.0	60.0	(60-95)
U_{H2} (%)	80.0	95.0	(50-95)
U_{O2} (%)	50.0	80.2	(25-95)
T_{EtOL} (ºC)	300.0	347.5	(200-400)
T_{eg} (ºC)	500.0	800.0	(200-800)

4. Conclusions

The development of an ethanol processor model within a modeling system as GAMS allowed the simultaneous flowsheet optimization and heat integration of the process. New operating conditions were obtained increasing the global efficiency of the fuel cell system about 5% higher than the reference case. Nevertheless due to the nonconvex characteristic of the NLP model, it is not possible to guarantee a global optimum, which will be addressed in further works. The results demonstrate the importance of optimal heat integration and the mathematical programming techniques for fuel cell-based processes.

References

M. Benito, J.L. Sanz, R. Isabel, R. Padilla, R. Arjona, L. Daza, 2005, Bio-ethanol steam reforming: Insights on the mechanism for hydrogen production, J. Power Sources,151,11-17.
M. A. Duran, I.E. Grossmann, 1986,Simultaneous optimization and heat integration of chemical processes, AIChE J.,32,123-138.
J.A. Francesconi, M.C. Mussati, R.O. Mato, P.A. Aguirre, 2007, Analysis of the energy efficiency of an integrated ethanol processor for PEM cuel cell systems, J. Power Sources, 167, 15-161.
J. Godat, F. Marechal, 2003, Optimization of a fuel cell system using process integration techniques, J. Power Sources, 118, 411-423.
J. Larminie, A. Dicks, 2000, Fuel Cell Systems Explained, p247-248.
C. Xu, L.T. Biegler, M.S. Jhon, 2006, Systematic optimization of an H_2 PEM fuel cell power generation system with heat integration, AIChE J., 52, 2496-2506.

Acknowledgements

1797

Reliability Modeling of a Natural Gas Recovery Plant Using *q*-Weibull Distribution

Isabel Sartori[a], Edilson M. de Assis[a], Adilton L. da Silva[b], Rosana L. F. Vieira de Melo[a], Ernesto P. Borges[b] e Silvio A. B. Vieira de Melo[a,*]

[a] Programa de Engenharia Industrial - PEI, Escola Politécnica, Universidade Federal da Bahia – UFBA, R. Aristides Novis, n. 2, Federação, Salvador - BA, 40210-630, Brasil
[b] Escola Politécnica, Universidade Federal da Bahia – UFBA, R. Aristides Novis, n. 2, Federação, Salvador - BA, 40210-630, Brasil
[*]E-mail: sabvm@ufba.br (corresponding author)

Abstract

We apply a recent generalization of Weibull distribution within the context of reliability engineering, inspired on mathematical developments that followed nonextensive statistical mechanics. Weibull distribution, largely used in reliability engineering, has been generalized to a q-Weibull distribution, following the lines of the q-exponential function, that appear within the context of nonextensve statistical mechanics. This new distribution is used to describe failure rate of a compression unit in a typical natural gas recovery plant currently in operation, based on time to failure data from historic process knowledge. The results show that the q-Weibull fits better the available data and this new distribution is very promising.

Keywords: reliability, natural gas, q-Weibull.

1. Introduction

A growing focus has been placed on Reliability, Availability, Maintainability and Safety (RAMS) during the design and operation of industrial systems, mainly due to the size and complexity of modern industrial plants, including oil and gas processing. Safety operation of natural gas production and recovery facilities depends on the continuous improvement of their reliability. A comprehensive assessment of operational safety requires a systemic approach based on statistical models for description of failure rates related to the equipments and their components. Development, choice or even application of a model to accurately characterize the failure rate is a non trivial task and mathematical simulation of reliability performance depends crucially on it. However, life data are still scarce and a predictive capability of the model should be often investigated. Reliability modeling, that is one of the most important steps for RAMS assessment, plays a key role to deal with incomplete life data..

There are many different statistical distributions that can be used to model life data. However, for some systems the classical life distributions are not satisfactory. Along the last two decades there has been a continuous and increasing development on what has been known as nonextensive statistical mechanics. The 1988 seminal paper by Tsallis (Tsallis, 1988) has introduced a generalization of the concept of entropy, by means of a real parameter q:

$$S_q = k \frac{\sum_i^w p_i^q - 1}{1-q} \tag{1}$$

where k is a positive constant that gives dimensional consistency to the expression (Boltzmann's constant), p_i is the probability of occupancy of the i-th microstate, and W is the total number of microstates of the system. This expression recovers the celebrated Boltzmann-Gibbs-Shannon (BGS) entropy, $S_1 = -k\sum_i^w p_i \ln p_i$ when $q \to 1$.

Nonextensive statistical mechanics has induced generalizations in other fields, e.g., in mathematics. A pair of functions naturally appears from the very beginning of the formalism, namely the q-logarithm, and its inverse, the q-exponential, defined as (Tsallis, 1994 and Borges, 1998)

$$\exp_q x = \left[1 + (1-q)x\right]_+^{1/(1-q)} \tag{2}$$

where $[x]_+ = \max\{x, 0\}$ and $q \in \square$. The q-exponential has many interesting properties see, for instance, Yamano (2002), as $\lim_{q \to 1} \exp_q x = e^x$, $\exp_q 0 = 1$ and

$$\int \exp_q x \, dx = \frac{1}{2-q} \exp_{1/(2-q)}\left[(2-q)x\right]. \tag{3}$$

Generalized statistical distributions developed within nonextensive statistical mechanics have been found applications in several physical, social and artificial complex systems (Gell-Mann and Tsallis, 2003). This q-type distributions (Nadarajah and Kotz, 2006, 2007) usually better fit data when compared with their equivalent "classical" versions (Picoli et al., 2003; de Souza and Tsallis, 1997), particularly if the tails are not well adjusted by Weibull distribution. Time-to-failure data has already been described by one of these generalizations, the q-Weibull distribution, by Costa et al. (2007), in the analysis of dielectric breakdown of ultra-thin oxides.

In this work we consider q-Weibull distribution to describe failure rate of a compression unit in a typical natural gas recovery plant currently in operation, based on time-to-failure data from historic process knowledge.

2. Reliability Modeling

Weibull distribution is largely used in reliability context. The probability density function for time to failure of the Weibull distribution in three-parameter form is defined by (Weibull, 1951)

$$f(t) = \frac{\beta(t-t_0)^{\beta-1}}{(\eta-t_0)^\beta} \exp\left(-\left(\frac{t-t_0}{\eta-t_0}\right)^\beta\right) \text{ with } t \geq t_0, \tag{4}$$

where t is time to failure or life time, β is the shape parameter, η is the characteristic life and t_0 is the location parameter.

Using the definition of the q-exponential (Eq. 2) for the generalization of Weibull distribution (Eq. 4), the q-Weibull probability density function can be written as

$$f_q(t) = \frac{\beta(2-q)}{\eta - t_0}\left(\frac{t-t_0}{\eta - t_0}\right)^{\beta-1} \exp_q\left(-\left(\frac{t-t_0}{\eta - t_0}\right)^{\beta}\right) \text{ with } t \geq t_0. \tag{5}$$

The reliability function is $R_q(t) = \int_t^{\infty} f_q(x)dx$, i.e. (See Eq. 3),

$$R_q(t) = \exp_{\frac{1}{2-q}}\left(-(2-q)\left(\frac{t-t_0}{\eta-t_0}\right)^{\beta}\right) \text{with } t \geq t_0. \tag{6}$$

Sample data are time-to-failure rank in ascending order and an estimate of the unreliability can be obtained using an approximation of the median ranks, also known as Benard's approximation, given by (Johnson, 1951)

$$\hat{F}_i = \frac{i-0.3}{n+0.4}, \tag{7}$$

where n is the sample size, i is the order number of failure varying from 1 to n. In this way, for every sampling time t_i we obtain

$$x_i = \ln(t_i - t_0) \tag{8}$$

$$y_i = \ln\left(-\ln_{\frac{1}{2-q}}\left(1-\hat{F}_i\right)\right). \tag{9}$$

This change of variables permits to estimate η parameter by least squares method for the linear correlation between x_i and y_i

$$\eta = \exp\left(\frac{-b}{\beta}\right)(2-q)^{\frac{1}{\beta}} + t_0. \tag{10}$$

Then, in this way, the q-Weibull distribution parameters are calculated searching the parameters q and t_0 that return the maximum value of R^2.

3. Results

Time to failure data from a compression unit of natural gas are gathered from historic industrial plant operation.

Results of q-Weibull fit to these data are evaluated through R^2 maximization ($R^2 = 0,986$ in Fig. 1) and the following parameters are found: $t_0 = 1.23h$, $q = 0.056$, $\beta = 0,502$, $\eta = 5576h$. Fig. 2 shows the data calculated with Eqs. 8 and 9 fitted to q-Weibull
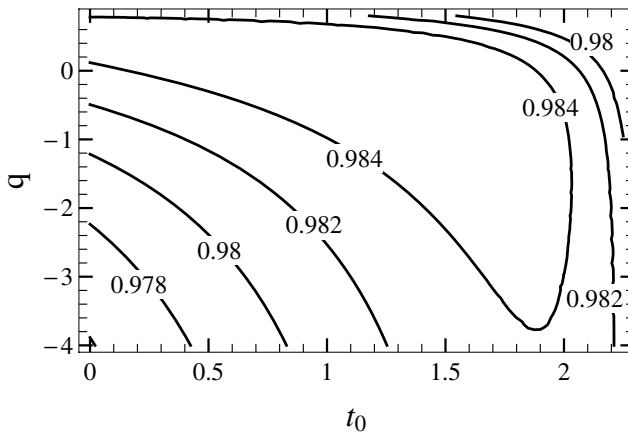
Fig.1 – Square correlation coefficient R^2 as a function of q and t_0. Lines indicate constant R^2, exhibiting the maximum in the 2-D surface.
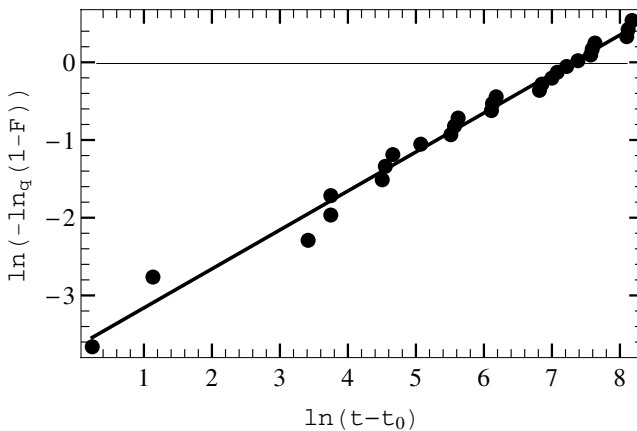
Fig. 2 – Failure probability as a function of shifted time-to-failure, properly scaled, in order to linearize the data. Square correlation coefficient is $R^2 = 0.986$.

For the classical Weibull distribution (Fig. 3) the following parameters are found: $t_0 = 0h$, $R^2 = 0,981$, $\beta = 0,609$ and $\eta = 826h$. One can observe that the values of η are quite different for both models and q-Weibull presents a bit higher value of R^2 coefficient. q-Weibull usually fits better than conventional Weibull if data are not adjusted well in the tail. This occurs because the q parameter unties the q-Weibull distribution of the exponential relation and the tail behaviour is different from conventional Weibull. Comparing the last three points in Fig.2 and 3, it can be seen that they are better fitted to the model in Fig. 2 where q-Weibull distribution is used.
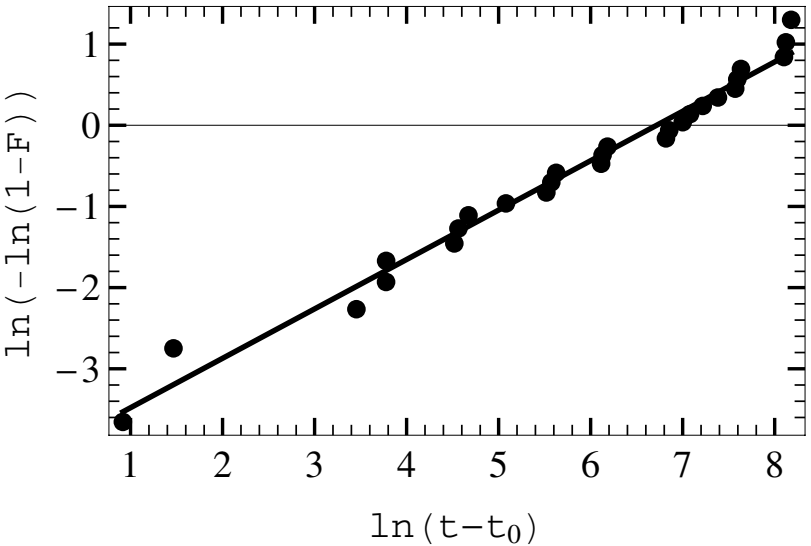
Fig. 3 – Failure probability as a function of shifted time-to-failure, properly scaled, in order to linearize the data. Square correlation coefficient is $R^2 = 0.981$ and parameter $q = 1$.

4. Conclusion

In this work a comparison is done between the Weibull classical distribution and the q-Weibull generalized distributions in the context of reliability engineering. Both distributions are applied to describe life data of a compression unit in a typical natural gas recovery plant. The results show that the q-Weibull distribution fits better to the available data and is very promising for reliability modeling of industrial cases. Further research is in progress to extend this methodology to other models and applications for reliability systems engineering.

References

E. P. Borges, 1998 , On a *q*-generalization of circular and hyperbolic functions, J. Phys. A: Math. Gen. 31, 5281-5288

U.M.S. Costa, V.N. Freire, L.C. Malacarne, R.S. Mendes, S. Picoli Jr., E.A. de Vasconcelos and E.F. da Silva Jr., 2006, An improved description of the dielectric breakdown in oxides based on a generalized Weibull distribution, Physica A, 361, 209-215

M. Gell-Mann and C. Tsallis (Eds.), 2003; Nonextensive Entropy – Interdisciplinary Applications. Oxford University Press, Oxford

L.G. Johnson, 1951, The Median Ranks of Sample Values in Their Population with an Application to Certain Fatigue Studies, Industry Mathematics, 2

S. Nadarajah and S. Kotz,2006, *q* -exponential is a Burr distribution, Phys. Lett. A 359, 577-579

S. Nadarajah and S. Kotz, 2007, On the *q*-type distributions, Physica A, 377, 465-468

S. Picoli Jr., R.S. Mendes and L.C. Malacarne, 2003, *q*-exponential, Weibull, and *q*-Weibull distributions: an empirical analysis, Physica A, 324, 678-688

A.M.C. de Souza and C. Tsallis, 1997, Students *t*- and *r*-distributions: Unified derivation from an entropic variational principle, Physica A 236, 52-57

C. Tsallis, 1988, Possible generalization of Boltzmann-Gibbs statistics, J. Stat. Phys., 52, 479

C. Tsallis, 1994, What are the numbers that experiments provide?, Química Nova, 17, 468-471

W. Weibull, 1951, A Statistical Distribution Function of Wide Applicability, Journal of Applied Mechanics-Transactions of the ASME, 18, 293-297

T. Yamano, 2002, Some properties of *q*-logarithm and *q*-exponential functions in Tsallis statistics, Physica A 305, 486-496

Simulation of Process Interesterification in Fluidized Bed Bioreactor for Production of Biodiesel

Jocélia S. Mendes, Jouciane S. Silva, Andrea L. O. Ferreira, Giovanilton F. Silva

Department of Chemical Engineering, UFC-Universidade Federa do Ceará
Campus do Pici, Bloco 709, Pici, 60455-760 – Ceará – Brazil

Abstract

The aim this work was investigated a new route for biodiesel production using immobilized lipase from *Candida antarctica* in continuous Fluidized Bed Bioreactor. Conventionally, Biodiesel (fatty acid methyl esters) is produced by transesterification in which, oil or fat is reacted with a monohydric alcohol in presence of a catalyst. In recent years, the use of lipases as biocatalysts for biodiesel production has become of great interest due to its environment friendly. But some alcohols such as methanol inactivated the lipases to some extent and the enzymatic stability was poor. In order to enhance the stability of the lipase, three-step methanolysis was adopted, however, glycerol, as one of the products was easy to adsorb on the surface of lipase resulting in serious negative effect on the enzymatic activity. For to solve problems was used the interesterification kinetics of triglycerides and methyl acetate for biodiesel production was modeled. A heterogeneous model describing the interesterification process in an immobilized enzyme fluidized-bed bioreactor was developed. A simplified model based on Ping Pong Bi Bi with substrate competitive inhibition mechanism was proposed to describe the reaction kinetics of the interesterification. The model without any adjustable parameters was used to predict the interesterification process. The key parameters which measured the extent of external and internal mass-transport resistances, as well as the degree of back-mixing were quantified and discussed. The fluidized-bed bioreactor considered in this investigation is composed of two phases: a fluid phase comprised mainly of the triglycerides and methyl acetate and the product (Biodiesel); and a solid phase which is the immobilized enzyme. The effects of some operating and design parameters on the performance of the fluidized-bed bioreactor were also analyzed. The model was also tested for its sensitivity to changes in hydrodynamic parameters.

Keywords: biodiesel, lipase from *Candida antarctica*, interesterification, fluidized bed bioreactor.

1. Introduction

Biodiesel is mono-alkyl esters of long chain fatty acids derived from renewable vegetable oils and animal fats. Biodiesel (fatty acid methyl esters) is produced by transesterification in which, oil or fat is reacted with a monohydric alcohol in presence of a catalyst. In the transesterification of vegetable oils, a triglyceride reacts with an alcohol in the presence of a strong acid or base, producing a mixture of fatty acids alkyl esters and glycerol. The overall process is a sequence of three consecutive and reversible reactions, in which monoglycerides are formed as intermediates. The stoichiometric reaction requires 1 mol of a triglyceride and 3 mol of the alcohol.

However, an excess of the alcohol is used to increase the yields of the alkyl esters and to allow its phase separation from the glycerol formed.

Biodiesel has become more attractive recently because of its environmental benefits and the fact that it is made from renewable resources (Schuchardt et al., 1998). The transesterification of vegetable oils with methanol or ethanol as well as the main uses of the fatty acid methyl esters are studied in this paper. The general aspects of this process and the applicability of different types of catalysts (acids, alkaline metal hydroxides, alkoxides and carbonates, enzymes and non-ionic bases, such as amines, amidines, guanidines and triamino(imino)phosphoranes) (Jordan and Gutsche, 2001; Fangrui and Hanna, 1998; Filippis et al., 1995; Antolín et al., 2002) are described, and ambient or elevated pressures and temperatures.

Conventionally, biodiesel was produced by transesterification of triglycerides and alcohols in the presence of an acid or an alkaline catalyst. In recent years, the use of lipases as biocatalysts for biodiesel production has become of great interest due to its environment friendly. However, some alcohols such as methanol inactivated the lipases to some extent and the enzymatic stability was poor. In order to enhance the stability of the lipase, three-step methanolysis was adopted, however, glycerol, as one of the products was easy to adsorb on the surface of lipase resulting in serious negative effect on the enzymatic activity Xu et al (2003), Dossat et la (1999) and Du et al (2004).

In order to solve the above-mentioned problems, we previously reported that using methyl acetate as acyl acceptor instead of methanol for biodiesel production could enhance the stability of the lipase significantly and in the process, triacetylglycerol instead of glycerol would be produced and it has been demonstrated that triacetylglycerol had no negative effect on the activity of the lipase. Moreover, triacetylglycerol was an important by-product with a higher value than glycerol and this novel route was thought to be very promising for large scale production of biodiesel.

2. Model Development

2.1. Kinetic Model

The enzymatic model used to describe the interesterification kinetics was based on the model Ping Pong Bi Bi with competitive inhibition for the substratum. The equation of that mechanism is shown:

$$V_i = \frac{V_{\max}[TG][AM]}{K_{M_{TG}}[AM]\left(1+\frac{[AM]}{K_I}\right)+K_{M_{AM}}[TG]+[TG][AM]} \tag{2.1}$$

where V_i was the initial reaction rate; $[TG]$ and $[AM]$ the initial molar concentrations of triglycerides and methyl acetate, respectively; K_{mTG} and K_{mA} the apparent Michaelis constants for triglycerides and methyl acetate, respectively; K_i the apparent inhibition constant of methyl acetate and Vmax the initial maximum velocity of the reaction. The kinetic constants are shown in

Table 1 The kinetic constants, Xu et al (2005).

Parameters	Value
$V_{\max}$ (mol/lmin)	1.9
K_{mTG} (mol/l)	1.0
K_{mA} (mol/l)	16
K_i (mol/l)	0.0455

2.2. Solid Phase Mass Balance

The fluidized-bed bioreactor considered in this investigation is composed of two phases: a fluid phase comprised mainly of the substrate (oil and methyl acetate) and the product (Biodiesel); and a solid phase which is the immobilized enzyme. The following assumptions are employed in the model: (1) the system is isothermal; (2) the movement of reactant within the biocatalyst can be described mathematically by Fick's law of diffusion where the effective diffusion coefficient is constant and independent of concentration; (3) the enzyme activity is uniform throughout the particle; (4) the fluid phase back-mixing can be quantified by an axial dispersion coefficient.

The general mass balance equation governing the concentration distribution in the fluid phase of the biocatalyst is given by:

$$D_L \frac{d^2 C_f}{dZ^2} - U_0 \frac{dC_f}{dZ} - (1-\varepsilon) r_V \qquad (2.2)$$

and the Danckwert's boundary conditions are:

$$\text{when } Z = 0, \quad -D_L \frac{dC_f}{dZ} + U C_f = U C_f \qquad (2.3)$$

$$\text{when } Z = L, \quad \frac{dC_{x,P,S}}{dZ} = 0 \qquad (2.4)$$

This set of boundary conditions, that is equations 7 and 8, is widely used in modeling immobilized enzyme fluidized-bed bioreactors.

2.3. Solid Phase Mass Balance

The general mass balance equation governing the concentration distribution in the fluid phase of the biocatalyst is given by equation 2.5.

$$\frac{dC_S}{dZ} = k_m (C_f - C_S) \qquad (2.5)$$

with the boundary conditions given by:

$$\text{when } Z = 0, \quad C_S = 0 \qquad (2.6)$$

2.4. Evaluation of the relevant parameters

The relevant parameters relating to the dispersion coefficient, external mass-transfer resistance and bed voidage are evaluated using correlations obtained from the literature and are given below.

2.4.1. Dispersion coefficient

The dispersion coefficient, D_L, in the fluidized-bed containing light, non-porous, solid particles was evaluated using the correlation of Tang and Fan (1990) given by Equation (2.7) below.

$$D_L = \frac{U_0 L}{Pe} \qquad (2.7)$$

$$\frac{1}{Pe} = 4.35 \left(\frac{\rho_s}{\rho_f} \right)^{2.64} \varepsilon^{1.467} \qquad (2.8)$$

The external particle-liquid mass-transfer coefficient was obtained using the correlation of Nore et al. (1992) given by:

$$k_m = 1.1 \left(\frac{U_0}{\varepsilon} \right)^{0.43} D_p^{0.24} \qquad (2.9)$$

The bed porosity in the bed is determined by the mechanics of fluidization. Hence, a realistic mathematical expression for the bed fluidization is necessary. For bed of uniform spherical particles, the following relation was proposed. The Richardson and Zaki (1954) relate bed voidage to the up-flow superficial liquid velocity given as:

$$\varepsilon = \left(\frac{U_0}{U_t} \right)^{1/n} \tag{2.10}$$

The terminal velocity given by Equation (2.11) was obtained using the correlation of Khan and Richardson (1990).

$$U_t = \frac{\mu}{\rho_f D_p} \left(2.33 Ga^{0.018} - 1.53 Ga^{-0.016} \right)^{13.3} \tag{2.11}$$

where

$$Ga = \frac{(\rho_s - \rho_f) \rho_f D_p^3 g}{\mu^2} \tag{2.12}$$

The fractional conversion, Y, shown in Fig. 1 against the flow rate is defined by Equation (2.13).

$$Y = \frac{C_0 - C_L}{V_{max} \, L/u} = \frac{X}{\theta} \tag{2.13}$$

where θ dimensionless residence time.

3. Results and Discussion

The modeling equations presented here constitute a system of non-linear boundary value problems which were solved by finite difference techniques. The details of solution methodology are omitted here for the sake of brevity.

Figure 1 shows the concentration of each composition triglycerides, methyl acetate and biodiesel against the height of the fluidized bed. From the Figure 2, it can be observed that, the model indicate the high concentration of biodiesel were reached in the end of the reactor. Indicating that the dispersion has influences in the interesterification of triglycerides for biodiesel production with methyl acetate.
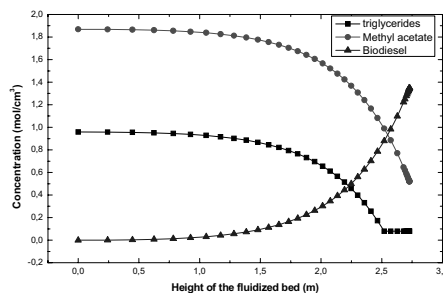
Figure 1 – The concentration of triglycerides, methyl acetate and biodiesel.

The Figure 2 show the conversions obtained for the model for fluidized bed in function of the dimensionless residence time. It can be observed that the measure that the dimensionless residence time increases the values of the conversions increased. Therefore, they are high necessary times of residence for the reactor to reach high conversions.
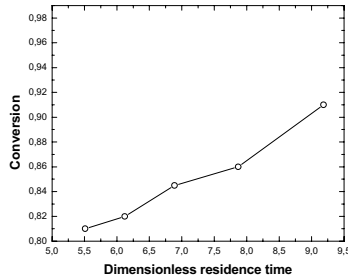
Figure 2 – Conversion in fluidized bed bioreactor

Another parameter of common interest in the design of fluidized bed reactors is the Peclet number which is the measure of the degree of back-mixing in the flow vessel. As the Peclet number approaches infinity plug-flow behavior is approached, while complete mixing is approached for a Peclet number approaching zero. It is obvious from the plot of Peclet number against the dimensionless residence time. The according Equations 2.7 and 2.8 when the value the Peclet increases the dispersion coefficient value decrease. This is also indicative of dominance of mixing over segregation. Thus, the conversion increases when the value de Peclet increases.
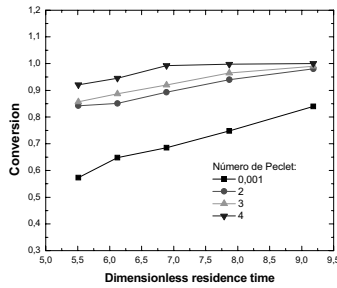
Figure 3 – Effect of number of Peclet on the conversion de the fluidized bed.

4. Conclusions

In this work a continuous fluidized bed bioreactor model with Immobilized Candida Antarctica in continuous Fluidized Bed Bioreactor was developed for simulating the steady-state performance of a bioreactor for production biodiesel. It was examined effect Peclet number and dimensionless residence time at conversion in fluidized bed. It the Peclet number decreases with the increase conversion production of biodiesel. It can be observed that conversion increased when the dimensionless residence time increased. The model can be used to simulate the biodiesel production by interesterification in fluidized bed bioreactor and the results of the simulation were satisfactory.

List of symbols

C_f	fluid phase concentration, mol/l	P_e	Peclet number, dimensionless
C_S	solid phase concentration, mol/l	u	flow rate
D_L	dispersion coefficient, m²/s	U_0	liquid superficial velocity, m/s
D_P	particle diameter, m	U_t	terminal velocity of particle
g	gravitational acceleration, m/s	ε	porosity of the bed, dimensionless
Ga	Galileo's number, dimensionless	ρ_f	density of the fluid, kg/m³

k_m	mass-transfer coefficient	ρ_s	mean particle density, kg/m^3
K_m	Michaelis constant	μ	viscosity, kg/m.s
L	reactor length, m	Y	Conversion

Subscripts

P	production concentration
S	substrate concentration
x	biomass concentration

References

Antolín, G., Tinaut, F. V., Briceño, Y. et alii (2002). Optimisation of biodiesel production by sunflower oil transesterification. Bioresource Technology, 83, 111-121.

Dossat, V.; Combes, D.; Marty, A. (1999) Lipase-catalysed transesterification of high oleic sunflower oil, Enz. Microb. Technol., 30 (1), 90-94.

Du, W.; Xu, Y. Y.; Liu, D. H.; Zeng, J. (2004) Comparative study on lipase-catalyzed transformation of soybean oil for biodiesel production with different acyl acceptors. Journal Molecular. Catalysis . B: Enzyme., 30 (3-4), 125-129.

Khan, A. R. & Richardson, J. F. (1990). Pressure gradient and friction factor for sedimentation and fluidization of uniform spheres in liquidg Chem. Engng. Sci., 45, 255-265.

Fangrui, M., Hanna, M. A. (1998). Biodiesel production: a review. Bioresource Tecnology, 70, 1-9.

Filippis, C. G. Scarsella, M., Sorrentino, M. (1995). Transesterification process for vegetable oils: A simple control method of methyl ester content. Journal American Oils Chemistry Society, 72, 1399-1407-1415.

Jordan, V., Gutsche, B. (2001). Development of environmentalyy benign processes for the production of fatty acid methyl esters. Chemosphere, 43, 99-108.

Nore, O., Briens, C., Margarilis, A. & Wild, G. (1992). Hydrodynamics, gas-liquid mass transfer and particle liquid heat and mass transfer in a three-phase fluidized bed for biochemical process application. Chem. Engng. Sci., 47, 3573-3580.

Richardson, J. F. & Zaki, W. N. (1954). Sedimentation and fluidization: Part I~ Trans. Inst. Chem. Engrs., 32, 35-53.

Schuchardt, U., Sercheli, R., Matheus, V. R. (1998). Transesterification of vegetable oils: a Review. Journal of the Brazilian Chemical Society, 9, 199.

Tang, W. T. & Fan, L. S. (1990). Axial liquid mixing in liquid-solid and gas-liquid fluidized-beds containing loq density particles. Chem. Engng. Sci., 45, 543-551.

Xu, Y. Y.; Du, W.; Liu, D. H.; Zeng J. (2003), A novel enzymatic route for biodiesel production from renewable oils in a solvent-free medium. Biotechnology letters, 25 (15), 1239-1241.Xu, Y. Y.; Du, W.; Liu, D. H. (2005), Study on the kinetic of enzymatic interesterification of triglycerides for biodiesel production with methyl acetate as the acceptor. Journal Molecular. Catalysis . B: Enzyme., 32, 241-245.

Kinetic Study of Biodiesel Production by Enzymatic Transesterification of Vegetable Oils

Fernando L.P. Pessoa[a], Shayane P. Magalhães[a] and Pedro W.C. Falcão[b]

[a]*DEQ - Department of Chemical Engineering, Chemistry School – Federal University of Rio de Janeiro, Av. Athos da Silveira Ramos, 149, Centro de Tecnologia, Bl. E, s. 201, Cidade Universitária, Ilha do Fundão, CEP 21941-909, Rio de Janeiro – RJ, Brazil*
[b]*PEQ/COPPE– Federal University of Rio de Janeiro, Av. Athos da Silveira Ramos, 149, Centro de Tecnologia, Bl. E, s. 201, Cidade Universitária, Ilha do Fundão, CEP 21941-909, Rio de Janeiro – RJ, Brazil*

Abstract

Although acid and alkaline catalysts are mostly used commercially for the synthesis of biodiesel, the difficulties related with the glycerol recovery, catalyst removal and the purification steps stimulate several studies of the enzymatic transesterification of oils since through this rout such problems are inexistent. Even though the high cost of the enzyme is still a drawback for the use of this process in industrial scale, the positive aspects of this technology motivate the researches in this field so that better operation conditions and competitiveness of the process are achieved. The Ping Pong Bi-Bi mechanism taking into account the competitive inhibition of the alcohol has been proposed to represent the kinetics of the enzymatic transesterification in many references in the literature, but the kinetic modeling of the enzymatic ethanolysis for the production of biodiesel has not been widely studied. In this work a kinetic modeling of the ethanolysis of sunflower oil catalyzed by lipase Novozym® 435 was carried out with kinetic parameters fitted to experimental data retrieved from the literature. The model was successful in predicting the experimental data with the reactor operating in a batch model. A semibatch reactor with continuous feed of ethanol was also investigated with respect to the influence of the ethanol flow rate on the inhibition of the enzyme.

Keywords: transesterification, lipase, ethanolysis, sunflower oil, biodiesel.

1. Introduction

The shortage of the fossil fuels and the environmental degradation due to gas emissions and/or spills have contributed for the increasing in the search for alternative fuels. Biodiesel is an alternative for the diesel derived from oil, which can be synthesized through the catalytic transesterification reaction of edible/non-edible oils or animal fats with an alcohol yielding fatty acid alkyl esters and glycerol as by product. It is environmentally friendly since it is derived from renewable and biodegradable resources and produces less gas emissions compared to the conventional fuels.

Several vegetable oils can be used as raw materials such as soybean, palm, sunflower, cotton, etc. Metanol and ethanol are the most used alcohols. In Brazil ethanol is advantageous due to its enormous availability. Acid, alkaline and biocatalysts can be employed commercially for the synthesis of biodiesel. The reaction is faster in the presence of alkaline catalysts, but when the concentration of water and fatty acids in the

oil is high the acid catalysts are preferable. As far as biocatalysts are concerned, lipases catalyze transesterification reactions of the triglycerides present in vegetable oils by acting on their ester bond. The difficulties during the glycerol recovery, catalyst removal and purification steps, which are characteristics of the reactions with alkaline and acid catalysts, are not found in transesterification routs that use immobilized lipases. Despite the high costs associated with the enzymes and their deactivation due to the presence of the alcohol, the positive aspects offered by this technology encourage the researches in this field in order to overcome the drawbacks so that the process can be used in industrial scale.

The schematic of the transesterification reaction for the production of biodiesel is shown in Figure 1. An excess of alcohol is usually required because the reaction is reversible.
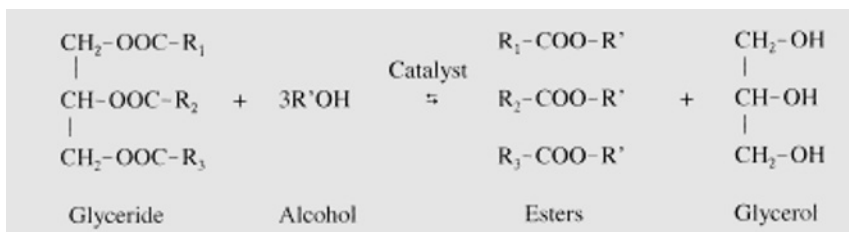
Figure 1. Scheme of triglycerides transesterification with alcohol.
R1, R2, R3, R' are the fatty acid chains

Kinetic data obtained experimentally can be modeled for scale-up purposes and also for the evaluation of the influence the different reaction parameters on the enzymatic transesterification reaction. Works that deal with modeling of the transesterification reaction for the production of biodiesel including the inhibition effect of the ethanol are seldom found in the literature. Therefore the objective of this work is to employ the Ping Pong Bi-Bi model to study the kinetic of the ethanolysis of sunflower oil with lipase Novozym$^{®}$ 435 as the catalyst.

2. Kinetic Modeling

The rate of an enzymatic reaction depends on factors such as substrates concentration, enzyme load, pH and temperature. Several studies in the literature report that the kinetics of the enzymatic transesterification reaction wherein raw materials derived from natural products are involved should be modeled using a model that accounts for the effects of the inhibition of the enzyme since some species present in the raw materials affect the activity of the enzyme. According to studies performed by Shimada *et al.* (2002), the deactivation of the enzyme is a consequence of the contact of the enzyme with the immiscible polar phase, containing the alcohol and glycerol, in the presence of the apolar oil phase. As reported by Dossat *et al.* (2002) e Al-Zuhair *et al*. (2006), a model widely used to describe the kinetics of this kind of reaction is based on the Ping Pong Bi-Bi mechanism which takes into consideration the influence of the alcohol and the substrates on the deactivation of the enzyme. In their model they considered the global reaction for the synthesis of biodiesel, as shown in Figure 1, instead of three consecutive reactions starting from tri-, di- and monoglyceride respectively because only small amounts of mono- and diglycerides were detected over the reaction. The rate expression given by Equation (1) was then used in the model.

$$Vi = \frac{V_m \cdot [oil] \cdot [alc]}{Km_{oil} \cdot [alc]\left(1 + \frac{[alc]}{Ki}\right) + Km_{alc}[oil] + [oil][alc]} \tag{1}$$

Where [oil] and [alc] are the molar concentrations of oil and alcohol, respectively. Km_{oil} and Km_{alc} are the biding constants for the oil and alcohol. K_i represents the inhibition constant of the alcohol, whereas V_m denotes the maximum initial rate for the reaction.

In this work a pseudo homogeneous kinetic model with the rate per unit of enzyme load, as used by Steinigeweg *et al*. (2004), combined with the Ping Pong Bi-Bi model was applied to describe the reaction rate as given by Equation (2).

$$(-r_i) = -\frac{1}{m_{cat}}\frac{1}{V}\frac{dNi}{dt} = \frac{V_m \cdot [oil] \cdot [alc]}{Km_{oil} \cdot [alc]\left(1 + \dfrac{[alc]}{Ki}\right) + Km_{alc}[oil] + [oil][alc]} \tag{2}$$

Where m_{cat} is the mass of enzyme, V is the reaction mixture volume and dNi/dt represents the number of moles of species *i* reacted per unit time.

2.1. Kinetic parameters

The experimental data taken from Hernández-Martín and Otero (2008), who studied the ethanolysis of sunflower oil catalyzed by lipase Novozym® 435, were used to fit the kinetic parameters (V_m, Km_{oil}, Km_{alc}, K_i). The reaction was carried out at ambient pressure, room temperature, initial molar ratio ethanol (pure) to oil of 20.6:1, enzyme load of 1g (50% w/w based on weight of oil) and using hexadecane as a solvent. Since the kinetic experimental data were referred as conversion of triglycerides, it was convenient to consider the sunflower oil as a pseudo component in the reaction model. Thus, the properties of the mixture of triglycerides and ethyl esters were calculated by averaging the individual properties with the corresponding mole fraction of their constituents.

The fitting of the parameters was performed with the applicative Mathcad v.14 by means of the solver *Minimize* employing the Quasi-Newton numerical method. The objective function (FO) is presented in Equation (3).

$$FO(V_m, Km_{oil}, Km_{alc}, Ki) = \sum_{j=1}^{n}\left[\frac{dNi}{dt}\bigg|_{calc\,j} - \frac{dNi}{dt}\bigg|_{exp\,j}\right]^2 \tag{3}$$

The results for the parameters in the model were: V_m=0.00024 mol.h^{-1}.mgcat^{-1}.L^{-1}, Km_{oil}=0.000006331 mol/L, Km_{alc}=0.04 mol/L, and K_i=0.00028797 mol/L. The kinetic parameters obtained maybe considered as apparent values due to the possibility of internal and/or external diffusion limitations.

3. Results and Discussion

In order to validate the kinetic parameters a batch reactor was modeled using a set of four ordinary differential equations representing the mole balances applied to the reactants and products, which was numerically solved with the *Odesolve* routine of Mathcad v.14 with the adaptive step-size algorithm that combines the fourth-order with fifth-order Runge-Kutta methods. The initial concentrations for the reactants, products and solvent were: oil=0.4603 mol/L, ethanol=9.47 mol/L, ester=0, glycerol=0, hexadecane=0.1780 mol/L. The initial reaction volume was 4.9616 cm^3. The comparison between the predicted and experimental conversions is illustrated in Figure 2. The modeling reproduced the experimental conversion data with an absolute percent mean deviation of 4.6%. According to this plot the model predicts rather successfully the experimental data and 90% conversion was achieved in approximately 4h.
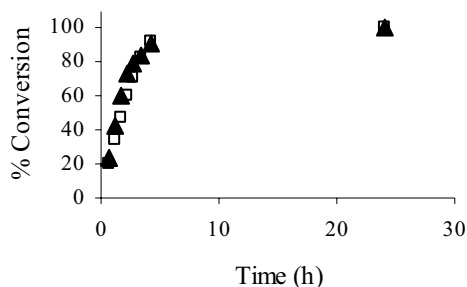
Figure 2. Sunflower oil conversion for a batch reactor. (▲) Exp., (□) Calc.
Conditions: 25°C, 1g enzyme (50% w/w), ethanol:oil molar ratio=20.6:1

The model was used to calculate the molar concentrations of reactants and products during the course of the reaction. The concentration profile for oil, ester and glycerol is presented in Figure 3. This plot shows that after 6 hours no changes occured in the concentrations indicating that the reaction is finished.
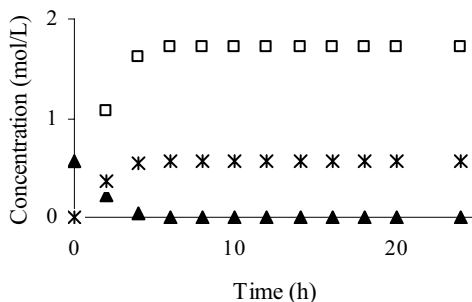
Figure 3. Concentration profiles for a batch reactor. (▲) oil, (□) ester, (*) glycerol
Conditions: 25°C, 1g enzyme (50% w/w), ethanol:oil molar ratio=20.6:1

Since some enzymes suffer from deactivation due to the presence of alcohols, the stepwise reaction has been mentioned in the literature (e.g. Watanabe *et al.* (1999) and Shimada *et al.* (2002)) as an option to overcome the inhibition due to high concentration of ethanol. In such a case the transesterification reaction is carried out with sequential addition of less than stoichiometric quantities of ethanol to a fixed amount of oil. Likewise, the use of a semibatch approach would be an alternative to conduct the reaction avoiding the deactivation of the enzyme.

Based on the kinetics discussed above, the modeling of a semibatch reactor with continuous feed of ethanol was conducted in order to investigate the influence of the concentration of ethanol on the conversion. Such a model applied to a semibatch reactor comprises a system of four ordinary differential equations representing mole balances on each species and an overall mass balance. The model is described by the following set of equations:

$$\frac{dN_{oil}}{dt} = \frac{-m_{cat}\left[V_m N_{oil}(t)\right]\left[\theta_{alc} - 3\left(1 - \frac{N_{oil}(t)}{N_{oil0}}\right)\right]}{Km_{oil}\left(\theta_{alc} - 3\left(1 - \frac{N_{oil}(t)}{N_{oil0}}\right)\right)\left(1 + \frac{N_{oil0}\left(\theta_{alc} - 3\left(1 - \frac{N_{oil}(t)}{N_{oil0}}\right)\right)}{VKi}\right) + Km_{alc}\left(\frac{N_{oil}(t)}{N_{oil0}}\right) + N_{oil}(t)\left[\theta_{alc} - 3\left(1 - \frac{N_{oil}(t)}{N_{oil0}}\right)\right]} \tag{4}$$

$$\frac{dN_{oil}}{dt} = \frac{1}{3}\frac{dN_{alc}}{dt} = -\frac{1}{3}\frac{dN_{ester}}{dt} = -\frac{dN_{glyc}}{dt} \tag{5}$$

$$\frac{dN_{alc}}{dt} = 3\frac{dN_{oil}}{dt} + \frac{v_0 N_{oil\,0}\theta_{alc}}{V_0} \tag{6}$$

$$V = V_0 + v_0 t \tag{7}$$

Where V_0=4.9616 cm^3, N_{oil0}=0.002284 mol and θ_{alc}=20.6 represent the initial conditions for the reaction mixture volume, moles of oil and ethanol to oil molar ratio, respectively. The mol balance for oil is given in Equation (4). Equation (5) gives the stoichiometry relating the reaction rates of each species (e.g. oil, alcohol, glycerol, and esters). Equation (6) shows the mol balance for ethanol including a continuous mol flow rate, whereas Equation (7) shows the volume as a function of time. Similarly to the case of the batch reactor, the model was solved with the *Odesolve* routine of Mathcad v.14.

In order to compare the conversions obtained with batch and semibatch strategies the initial conditions of the reaction mixture volume and concentrations of the reactants have the same values in both modes. The goal was to investigate the behaviour of the conversion by varying the flow rate of ethanol added to the system continuously, v_0. The effect of ethanol flow rate on the conversion in a semibatch reactor is illustrated in Figure 4, which shows the conversions over a period of 6h for ethanol flow rates of 0.1ml/h, 1ml/h and 10 ml/h. The improvement in the conversion was basically due to the increasing in the ethanol flow rate that yields high ethanol to oil molar ratios. It appears that, under the given operation conditions, the inhibition of the alcohol is not relevant. These results are in compliance with the findings of Hernández-Martín and Otero (2008) who observed rather insignificant deactivation of Novozym® 435 by ethanol with alcohol to oil molar ratios ranging from 14.3:1 to 28:8. Additionally, Irimescu *et al.* (2002) reported a satisfactory performance of this enzyme in the ethanolysis of triacylglycerol using even high molar ratio (e.g. 77:1). The increasing in the conversion caused by the reduction in the inhibition effects might be better evaluated by starting the reaction with a concentration of ethanol lower than that used in the batch mode and consequently ethanol to oil molar ratios lower that 20.6:1. However there was no experimental data available so that the kinetic parameters could be determined.
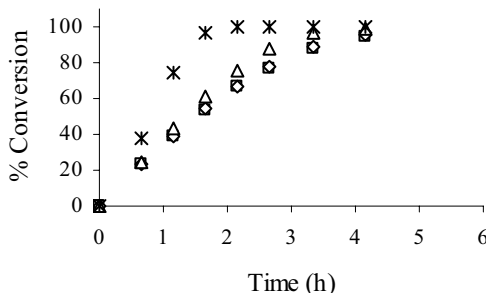
Figure 4. Conversion results for batch and semi batch reactors. (□) batch mode, (◊) semibatch with 0.1ml/h, (Δ) semibatch with 1.0 ml/h, (*) semibatch with 10 ml/h. Conditions: 25°C, 1g enzyme (50% w/w), initial molar ethanol:oil ratio=20.6:1.

4. Conclusion

A model based on the Ping Pong Bi-Bi mechanism with competitive ethanol inhibition was employed to describe the kinetics of the transesterification reaction of sunflower oil with ethanol catalyzed by lipase Novozym® 435. The kinetic constants were fit to

experimental data and the model was found to be qualitatively consistent with the proposed mechanism since it reproduced the experimental conversions rather adequately. The results of the model for a semibatch reactor with continuous addition of ethanol using the same initial conditions as in the batch mode showed improvements in the conversion basically due to the increasing in the ethanol to oil molar ratio rather than the decreasing in the inhibition effects.

5. Nomenclature

[alc]	molar concentration of alcohol, M
dNi/dt	number of moles of species i reacted per unit time, mol/h
$dNi/dt|_{exp}$	experimental number of moles of species i reacted per unit time, mol/h
$dNi/dt|_{calc}$	calculated number of moles of species i reacted per unit time, mol/h
Km_{oil}	biding constant for the oil, M
Km_{alc}	biding constant for the alcohol, M
K_i	inhibition constant of the alcohol, M
m_{cat}	mass of enzyme, mg
N_i	number of moles of species i, mol
N_{i0}	initial number of moles of species i, mol
[oil]	molar concentration of oil, M
R and R'	fatty acid chain
t	reaction time, h
V	reaction mixture volume, L
V_i	initial rate of reaction, $mol.h^{-1}.L^{-1}.mg^{-1}$
V_m	maximum initial rate, $mol.h^{-1}.L^{-1}.mg^{-1}$
V_0	initial reaction mixture volume, L
v_0	flow rate of ethanol, $L.h^{-1}$
θ_{alc}	initial molar ratio of ethanol to oil

References

S. Al-Zuhair, K.V. Jayaraman, S. Krishnan, W. Hoong Chan, 2006, The Effect of Acid Concentration and Water Content on the Production of Biodiesel by Lipase, Biochemical Engineering Journal, 30, 212-217.

V. Dossat, D. Combes, A. Marty, 2002, Lipase-Catalysed Tranesterfication of High Oleic Sunflower Oil, Enzyme and Microbial technology, 30, 90-94.

E. Hernández-Martín, C. Otero, 2008, Different Enzyme Requirements for the Synthesis of Biodiesel: Novozym® 435 and Lipozyme® TM IM. Biores. Technology, 99, 277-286.

R. Irimescu, Y. Iwasaki, C.T. Hou, 2002, Study of TAG ethanolysis to 2-MAG by immobilized *Candida Antarctica* lipase and synthesis of symmetrically structured TAG, J. Am. Chem. Soc. , 79, 879-883.

Y. Shimada, Y. Watanabe, A. Sugihara, Y. Tominaga, 2002, Enzymatic Alcoholysis for Biodiesel Fuel Production and Application of the Reaction to Oil Processing, Journal of Molecular Catalysis B: Enzymatic, 17, 133-142.

S. Steinigeweg, J. Gmehling, 2004, Transesterification Processes by Combination of Reactive Distillation and Pervaporation, Chem. Eng. and Processing, 43, 447-456.

Y. Watanabe, Y. Shimada, A. Sugihara, Tominaga, Y., 1999, Stepwise ethanolysis of una oil using immobilized *Candida Antarctica* lipase, J. Biosci. Bioeng., 88, 622-626.

1815

How Modeling can Help to Discriminate Assumptions on the Influence of Nitrogen Consumption on pH during Fermentation

Huberson AKIN[a], Cédric BRANDAM[a], Xuân-Mi MEYER[a], Pierre STREHAIANO[a]

[a] *Université de Toulouse-INPT-ENSIACET, Laboratoire de Génie Chimique, 118 route de Narbonne, 31077 Toulouse cedex 4, France*

Abstract

This work deals with the use of a mathematical model to investigate the influence of the nitrogenous source on the pH during alcoholic fermentation in winemaking. The application of the model to fermentation medium whose nitrogen source was made up only of ammoniac confirms the assumption that the assimilation of one mole of ammonia releases one mole of proton in the medium. The use of the model made it possible to invalidate two assumptions concerning the impact of the assimilation of the amino acids on the pH. The most probable assumption is that the assimilation of the molecules of amino acids charged positively led to the emission of protons in the extra cellular medium. The model including this assumption was also used successfully to predict pH in the case of a fermentation realised with natural grape must.

Keywords: nitrogen consumption, modelling, pH, fermentation

1. Introduction

In a recent context of an international competition, enologists seek to better understand the wine making phenomena to better control the production quality. If the source of carbon in the process of wine making comes primarily from the glucose and fructose available in the grape, the source of nitrogen is much varied and can take forms more or less complex: ammonium ions, amino acids, peptides and proteins.

Nitrogen compounds are quantitatively the second nutriment of the yeast, after the carbon element. It is known that only ammonium ions and some amino acids can be metabolised by yeasts and that the kinetic of the fermentation and aroma production are particularly influenced by the quantity and the nature of amino acids.

Various studies have reported a pH decrease during alcoholic fermentation by *Saccharomyces cerevisiae* on medium where ammonium ions were the only nitrogenous source (Kotyk, 1989, Siegler et al., 1981). According to Won et al. (1993) and Castrillo et al. (1995), when one mole of ammonium is consumed, one mole of proton is excreted resulting in a pH decrease. The mechanisms of assimilation of amino acids by yeast seem more complex. The type of exchanges taking place when amino acids pass through cell membrane is not yet known. Several assumptions on the relation between pH values and amino-acid assimilation can be drawn from these studies but until now, no work has permitted to discriminate them.

In a previous study, Akin et al. (2008) have proposed a mathematical model to calculate the pH of a grape must during the fermentation. This model has been validated on synthetic medium and natural grape musts.

The objective of this paper is to investigate the possibility to correlate the evolution of the pH in the fermentation media to the most probable mechanism of assimilation of amino-acids by the mean of a mathematical model. Three assumptions for the effect on pH of assimilation of amino acids were tested:

- Assumption 1: assimilation of amino acid does not influence the pH
- Assumption 2: consumption of one mole of amino acid causes the excretion of one mole of H^+, like for the ammonium consumption
- Assumption 3: consumption of one mole of amino acid charged positively only causes excretion of one mole of protons to respect the electroneutrality principle.

These different assumptions drawn on the assimilation of ammonium ions and amino acids are introduced in the model and experimental and simulated values are then compared to discriminate them.

2. Materials and methods

Saccharomyces cerevisiae QA-23 commercialised by Lallemand inc. was used as it is a classical yeast strain for white winemaking.

Fermentations were carried out on synthetic media whose composition was close to white grape must: glucose (200 g.L^{-1}), malic acid (6 g.L^{-1}), citric acid (6 g.L^{-1}), KH_2PO_4 (0.75 g.L^{-1}), K_2SO_4 (0.5 g.L^{-1}), $MgSO_4$, 7 H_2O (0.25 g.L^{-1}), $CaCl_2$, $2H_2O$ (0.16 g.L^{-1}), NaCl (0.2 g.L^{-1}), a mixture of oligo elements, vitamins and anaerobia factors. The nitrogen source differed with experiments: ammonium ions for an equivalent of 420 mg N.L^{-1} for the medium 1 and a mixture of amino acids for an equivalent of 384 mg N.L^{-1} for the medium 2. In this mixture, only 305 mgN.L^{-1} could be assimilated by yeasts since 79 mgN.L^{-1} was proline which can not be metabolised by *Saccharomyces cerevisiae*. The pH of media was adjusted to 3.3 before autoclaving with a sodium hydroxide solution (8N).

Experiments were also performed on a natural grape must provided by SOPAGLY Company. The grape must contains an equivalent of 74.9 mgN.L^{-1} of ammonium ions and an equivalent of 136 mgN.L^{-1} of a mixture of amino acids (114.8 mgN.L^{-1} could be assimilated by yeast, 21.2 mgN.L^{-1} was proline). Fermentations were carried out in New Brunswick Scientific (NBS) fermentor type of 2 litres at 20°C under a 150 rpm agitation. Temperature was regulated at 20°C. The pH was not controlled during this anaerobic fermentation. For all fermentations, the initial cellular concentration was fixed to 8.10^6 cells.mL^{-1}.

The cell concentration was estimated using an electronic analyser Beckman Coulter. Concentrations of ammonium, malic and lactic acids were measured by enzymatic kits (MicroDom). Citric, acetic and succinic acid, ethanol and glycerol were measured by HPLC method. Mineral elements (calcium ion, sodium ion, magnesium ion, potassium ion) were measured with an emission spectrometer of Horiba Jobin Yvon. Their concentration was evaluated on the initial medium. They were considered as constant during the fermentation. Individual amino acid concentrations were determined by Biochrom 30 method. The pH was measured using an external pH-meter (PHM210 Radiometer analytical) with a precision of 0.1 pH unit.

3. Mathematical model for pH calculation

The pH calculation model used in this study was based on the thermodynamic equilibrium of electrolytic compounds in solution. The molality of hydrogen ions and so the pH, are determined by solving a non linear algebraic equation system. For more details and model validations, it is possible to refer to Akin et al. (2008)

Instead of writing classically the mass balance on each species in solution, the mass balance was written on an invariant of the reactive system. Then, it is possible to express molalities of different amino acid and acids species (m_i) as an explicit function of m_H^+. The equilibrium dissociation constants for organic acids are calculated as a second order polynomial function of volumic percent in ethanol of the solution (Usseglio-Tomasset and Bosia, 1978). The pKa of NH_4^+/NH_3 and of all the amino acids were taken constant with ethanol concentration.

The activity coefficient of H^+ (γ_{H^+}) necessary to calculate pH according to the equation $pH = -\log_{10}(a_{H^+}) = -\log_{10}(\gamma_{H^+} m_{H^+})$ was calculated with a Debbye-Hückel model with the MacInnes convention. The calculation of the average coefficient of activity was carried out with the law of Debye-Hückel, in the scale of the molalities

New correlations for the density and the dielectric constant were established taking into account the influence of the sugar (0 to 200 g.L^{-1}) and the ethanol (0 to 12 % vol) at 20°C. The formulation of the model with explicit calculations of all the species in solutions as a function of the molality in H^+ and the ionic force I, enables to solve a non linear equation system consisting of only two equations: the ionic force equation and the electro neutrality equation, with two unknowns: the molality in H^+ and the ionic force I.

$$I = \frac{1}{2}\left[\sum_{i=1}^{n\_species} m_i z_i^2\right] \text{ where } z_i \text{ represents the charge of the electrolyte i}$$

$$\sum_{i=1}^{n\_electrolytes} m_i(m_{H^+}, I) \, z_i = 0$$

This formulation greatly facilitates the initialisation of the Newton-Raphson iterative procedure carried out to solve the problem.

4. Results and discussions

4.1. Fermentation of the medium 1

During the first fermentation period (0 to 30 hours), there was no biologic activity. During the second period, between 30 and 110 hours, the metabolic activity of yeast increased: a consumption of sugars and ammonium and a production of biomass and ethanol can be observed. At the end of the fermentation, ethanol concentration was about 82 g.L^{-1}, biomass attained almost 160.10^6 cells.mL^{-1} and ammonium was entirely consumed. The decrease of pH was also emphasized since its final value was 2.90.

The measurement of organic acid concentrations indicated that very small quantities were produced: 0.16 g.L^{-1} of succinic acid, 0.18 g.L^{-1} for acetic acid. Citric and malic acids remained constant to 6 g.L^{-1} during the whole fermentation.

The pH calculated with the model (3.37) at the very beginning of the fermentation when the medium composition was perfectly known, was in agreement with the experimental value. Ethanol, sugar, organic acids and ammonium concentrations have been measured at different times for which pH values can be calculated by the model.

For this medium, the assumption of one mole of proton released for each ammonium ion (in cationic form) consumed by *Saccharomyces cerevisiae* was in agreement with the experimental data (figure 1).
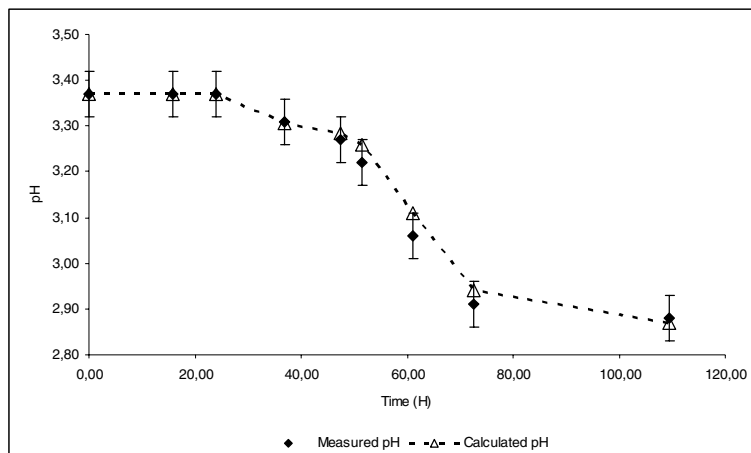
Figure 1: Comparison of experimental and simulated pH values in medium 1 (T=20°C)

4.2. Fermentation of the medium 2

During the first period of fermentation (0 to 20 hours), yeasts were in a lag phase and no changes were observed in the medium. During the second period (20 to 110 hours), ethanol production occurred to reach 94 g.L^{-1} at the end of the experiment. Concentration of biomass rose to 200.10^6 cellsml^{-1}and amino acids were totally consumed (except the proline). 0.3 g.L^{-1} of succinic acid and 0.3 g.L^{-1} of acetic acid were produced. Malic and citric acids remained constant to their initial concentration at 6 g.L^{-1}. The pH value decreased from 3.23 to 3.13 after 50 hours of the fermentation which corresponds to the end of the assimilation of amino acids. Afterwards, pH increased regularly to reach 3.25 at the end of the experiment. This rise was previously shown to be related to the influence of ethanol on the organic acids dissociation equilibriums. The pH evolution was simulated and compared to experimental data, under the three assumptions (figure 2).

Assumption 1:
Since organic acids production was very low, no other phenomenon except amino acid consumption could explain pH decrease. Model calculation confirmed this fact, by calculating a constant pH value for this assumption.

Assumption 2:
The minimum pH value calculated was 3.03, that was lower than the experimental one (3.13). The production of proton is then overestimated by the model. This would indicate that all amino-acids can not be considered at the same level. The assimilation of some of them may lead to the excretion of one or more protons whereas other may be without any effect on pH.

Assumption 3:

To test the third assumption, the total molalities of amino acids on cationic, anionic and neutral form during the fermentation were determined. The amino acids repartition in the medium, consisted initially of 70% amino acids globally under the neutral form, 27 % with one positive charge, 1.3 % with two positive charges and 1.7 % with one negative charge. A good adequacy between the calculated and experimental pH values was found (figure 2). The maximal deviation was 0.04 point of pH that is less than the precision of the measure.

The simulation results were globally in agreement with the experimental data when the third assumption was formulated.
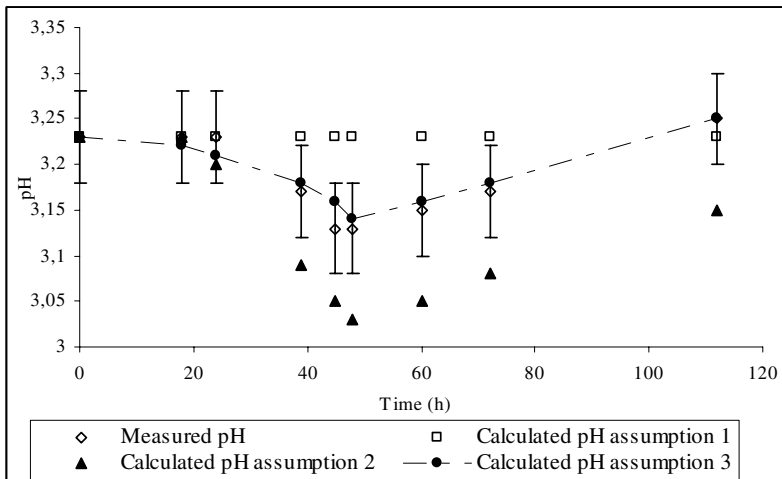
Figure 2: Comparison of experimental and calculated pH with assumption 1, 2 and 3 in MS_AA medium (Initial pH=3.23; T=20°C)

4.3. Fermentation of the natural white grape must

To confirm the assumption made for the effect on pH of assimilation of amino acids, fermentation was realised on a natural white grape must. The pH evolution was very similar to the synthetic media: a pH decrease (from 3.32 to 3.14) during the first 50 h of the fermentation corresponding to the concomitant ammonium and amino acid consumption, then a pH increase (from 3.14 to 3.3) corresponding to the ethanol production (88 g.L^{-1}). Organic acid production was more important than in synthetic media. There was a production of 0.7 and 0.5 g.L^{-1} of succinic acid and acetic acid. Lactic acid was not produced and citric acid remained constant to 1.5 g.L^{-1}. Malic acid was little consumed from 6 to 5.4 g.L^{-1}.

The quantities of amino acid on anionic, cationic and neutral forms are calculated for the grape must. At the beginning of the fermentation, it consists of 61% of amino acids in neutral form, 36 % with one positive charge, 1.8 % with 2 positive charges and 1.2 % with one negative charge. pH values were calculated assuming the third assumption are here again in a good agreement with experimental values (figure 3).
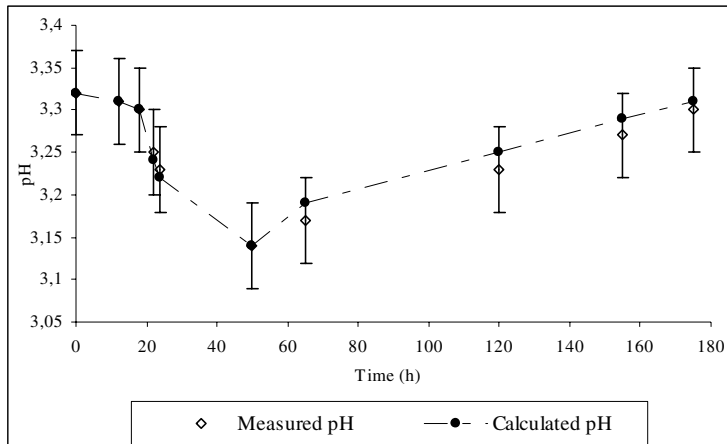
Figure 3: Comparison of experimental and calculated pH in grape must (Initial pH=3.32; T=20°C)

5. Conclusion

This study confirms that the acidification of the medium during alcoholic fermentation is directly linked to the assimilation of the nitrogen source. On a medium where nitrogen was only brought by ammonium (420 mg N.L^{-1}), the pH decreases down to 0.47 pH unit. With amino acids as nitrogen source (305 mg N.L^{-1} of assimilable nitrogen) the decrease of the pH was lower, only 0.1 unit of pH. The results of the pH calculation model have confirmed that the mechanism of assimilation of each ammonium ions by yeast releases one proton in the medium. For amino acids assimilation, several assumptions has been tested since literature was equivocal on the effect on the pH. The model allowed us to suggest that only the cationic form of amino acids was responsible for the pH decrease. This assumption has been tested successfully on a synthetic medium but also on a natural grape must. It must be validated with additional fermentations realised on medium with different ratio of amino acids.

References

A.Kotyk, Proton extrusion in yeast, In Fleischer S. and Fleischer B. (Eds) Method in enzymology, 174 (Biomembranes, part U) Academic Press, New York (1989) 592-603

K.Sigler, A.Knotkova and A.Kotyk Factor governing substrate induced generation and extrusion of proton in yeast. Biochim. Biophys. Acta 643 (1981) 572-582

J.I Castrillo., I.De Miguel, U.O.Ugalde, Proton production and consumption pathways in yeasts metabolism. A chemostat culture analysis, Yeast 11 (1995) 1353-1365.

J.I.Won, Y.L.Yang., B.G.Kim and C.Y.Choi., Adaptative control of specific growth rate based on proton production in anaerobic fed-batch culture. Biotechnology letters 15 (5) (1993) 511-516.

H.Akin, C.Brandam, X-M. Meyer and P. Strehaiano, A model for pH determination during alcoholic fermentation of a grape must by *Saccharomyces cerevisiae*, Chem. Eng. Process 47 (2008) 1986-1993

L.Usseglio-Tomasset and P.D.Bosia**,** Determinazione delle costanti di dissociatione dei principali acidi del vino in soluzioni idroalcoliche di interesse enologico, Rivisto di viticoltura e di enologia di conegliano 31 (1978) 380-403

Integrated Procurement and Operational Planning of a Biorefinery Considering Contracts and Futures

Choamun Yun,[a] Young Kim,[b] Jeongho Park,[a] Sunwon Park [a, *]

[a]*Department of Chemical and Biomolecular Engineering, KAIST, 335 Gwahangno, Daejeon 305-701, Korea*
[b]*Energy Plant Research Division, KIMM, 104 Sinseongno, Daejeon 305-343, Korea*

Abstract

The profit of a biorefinery is highly affected by the supply of its raw materials and margin from the product. Taking responsive actions to the unstable supply of raw materials and the fluctuating prices are of major concern for the efficient management of biorefineries. In this study, a biorefinery complex is defined to tackle these issues by diversifying products as well as raw materials. Various raw materials are purchased based on the contracts and the spot prices of candidate raw materials while their total amount depends on product demands. The supply of required raw materials is accompanied by the futures contracts to curtail the risks involved in the procurement. In the downstream process, operational planning of fermentation and separation units is established according to the demand and margin of multiple products. The proposed model of integrated planning for a biorefinery would contribute to escalating its profitability and operational flexibility.

Keywords: Integrated biorefinery; multiproducts; procurement planning; futures

1. Introduction

Biorefinery has been appreciated for its environmentally friendly process and products. Despite of the subsidies and tax advantages from many governments, however, the industry has not been grown sufficiently for its relatively low profitability (Lakhdar and Papageorgiou, 2008). Accordingly, developing an economically sustainable as well as commercially viable process has been a crucial issue. Different approaches on alternative media, enzymes, and processes with various reaction conditions have been studied (Porro et al., 1999; Steffens et al., 2000; Vishnu et al., 2000; Willke and Vorlop, 2001; Li et al., 2006). Koutinas et al. (2004; 2007) studied the alternative upstream processes for the production of generic microbial feedstocks while downstream processing options are investigated by Joglekar et al. (2006). The idea of producing multiple products from a biorefinery as in the petrochemical industry has also been explored by several studies (Tran et al., 2004; Kamm and Kamm, 2007; Sadhukhan et al., 2008; Sammons et al., 2008). After the desirable process is selected, profit of a biorefinery still varies with the operation of the process. As in other chemical processes, the productivity is uncertain due to the changing operation parameters such as the yield of batch and the purity after separation processes (Lim et al., 2005; Lakhdar and Papageorgiou, 2008). The more significant factors for the high profit variability are the fluctuating costs for raw materials and the price change of products (Franceschin, 2008). The prices of raw materials change for a variety of reasons such as seasonal

---

[*] *To whom correspondence should be addressed. E-Mail: sunwon@kaist.ac.kr*

effects, states of harvest, or policy changes; and the product prices change due to the changing market conditions.

One way of curtailing the risks is to diversify the raw materials and products. By varying the source of starch and the targeting markets, a biorefinery may be actively respond to the changes. On the other hand, the risk can also be reduced with financial derivatives such as options and futures (Yun et al., 2009; Park et al., 2009). In this work, a procurement and operational planning model is proposed to decrease the profit variability of a biorefinery by flexibly operating an integrated production process for multiple products. Moreover, the risks involved in the procurement are reduced by purchasing the diversified raw materials and their futures contracts.

2. Problem description

A biorefinery process manufacturing multiple products from diversified raw materials starts from milling of purchased raw materials, e.g., corn or wheat. The processed raw materials are liquefied and saccharified in the form of starch. Subsequently, solution containing glucose is generated from the saccharification unit and it is fed into fermenters with enzymes and additional media required for the production of a desired product. The products are manufactured from the shared operating units.

The outlet streams from fermenters are separated into two main downstreams: one stream for the production of ethanol, and the other for that of biochemicals. The broth fermented for ethanol production undergoes the ethanol recovery process consisting of a beer column and a rectifier. Other biochemicals of lactic acid, itaconic acid and citric acid are assumed to be purified through an identical procedure. The broths are filtered successively in a filtration unit, an ion exchanger and a decolorization unit. The remaining liquid is condensed before the products are crystallized and dried.

3. Operational planning of an integrated process

The main assumptions of the procurement and operational planning model are as follows: (a) Processing times are constant for every batch; (b) Clean-up times are ignored; (c) Sufficient raw materials are always available; (d) The operating costs of the units are constant.

 The profit to be maximized in this model is:

$$\sum_n Z_n = \sum_n \sum_k P_{k,n} \cdot X_{k,n} - \sum_n \sum_r P_{r,n} \cdot X_{r,n} - \sum_n \sum_j OC_j \delta_{j,n} - \sum_n \sum_i IC_i I_{i,n} \qquad (1)$$

Z_n is the profit of time period n, the revenue subtracted by raw material costs, operating costs, and inventory costs. $P_{k,n}$ and $X_{k,n}$ are the price and sales amount of product k; $P_{r,n}$ and $X_{r,n}$, the price and purchase amount of raw material r; OC_j, the operating costs of unit j; and IC_i, the inventory cost of storage tank i whose level is $I_{i,n}$. $\delta_{j,n}$ is a binary variable indicating whether operating unit j starts its operation at the end of nth time period.

Inventory level of a raw material is increased with its procurement and decreased when there is an inlet stream to a milling unit, M_n^α, i.e.,

$$I_{r,n}^M = \sum_r (X_{r,n}^L + X_{r,n}^S) - M_n^\alpha \qquad (2)$$

where $X_{r,n}^L$ and $X_{r,n}^S$ are the purchasing amount of raw material r at time period n through long term and spot contracts, respectively.

Inventory levels between milling and liquefaction unit is balanced with the outlet stream from the milling unit, $M_{r,n}^{\beta}$, and inlet to the liquefaction unit, V_l^{α}.

$$I_n = I_{n-1} + M_n^{\beta} - V_l^{\alpha}\delta_{l,n} \tag{3}$$

Similarly, intermediate storage level, $I_{i,n}$, between adjacent units j and j' are calculated as below.

$$I_{i,n} = I_{i,n-1} + V_j^{\beta}e_{j,n} - V_{j'}^{\alpha}\delta_{j',n} \qquad \text{for } \forall n \tag{4}$$

$e_{j,n}$ defines whether operating unit j ends its operation at the end of nth time period. In the downstream process from fermentor units,

$$I_{i,k,n} = I_{i,k,n-1} + V_j^{\beta}e_{j,k,n} - V_{j'}^{\alpha}\delta_{j',k,n} \qquad \text{for } \forall k,n \tag{5}$$

For final storages, the outlet stream is the sales amount of product k, $X_{k,n}$.

$$I_{i,n} = I_{i,n-1} + V_j^{\beta}e_{j,n} - X_{k,n} \qquad \text{for } \forall k,n \tag{6}$$

Additionally, the following time constraints are introduced to prevent the overlapped production plan in a unit.

$$e_{j,n+\tau_j} = \delta_{j,n} \tag{7}$$

$$e_{j,n} = 0 \quad \text{for } n \cdot \Delta t \le \tau_j \tag{8}$$

$$\sum_{u=n}^{n+\tau_j} \delta_{j,u} \le 1 \tag{9}$$

where τ_j is the processing time at unit j and Δt is the length of time period n.

4. Risk management for procurement

The prices of the raw materials are not necessarily cheaper or the same as the expectation at the point of purchase. Accordingly, the profit of a biorefinery may alter with the changing prices to a great extent even when the plant is operated following the optimal procurement and operational plans. In this study, futures contracts are considered for hedging the risks involved in the purchase of raw materials.
The profit generated from the futures transaction is as below.

$$\sum_r (PF_{r,1} - PF_{r,0})XF_r \tag{10}$$

where $PF_{r,0}$ and $PF_{r,1}$ are the futures prices when taking the position at time period $n = 0$ and closing out the position at time period $n = 1$, respectively. The size of futures contracts, XF_r, is decided depending on the hedging strategy. Here, the objective is to hedge against changes in the price of raw materials. As the spot and futures contracts are generally purchased independently, it is rational to add the profits generated by the both. Consequently, the profit of a biorefinery is redefined as

$$\sum_{t} Z_t + \sum_{t} \sum_{r} (PF_{r,1} - PF_{r,0}) XF_r \qquad\qquad (11)$$

5. Scenario generation

Planning models are calculated using the scenarios generated either by decision makers or by mathematical models based on historical data. The scenarios of commodity prices are generated by assuming that the price follows Geometric Brownian Motion (Schwartz, 1997). The scenarios of futures prices are generated by assuming that their deviation from spot prices follows a normal distribution.

6. Illustration

The developed model is illustrated with a simplified example process manufacturing ethanol, lactic acid, and itaconic acid, from corn and wheat.

As a first step, price scenarios are generated for raw materials, futures of raw materials, and products on the basis of historical data from National Agricultural Statistics Service (2008), Tran et al. (2004), and Kenyon (2001). Further pertinent data are provided in Yun et al. (2009).

The results obtained for the generated scenarios are shown in Tables 1 through 3. The flexible operational planning results in the slightly higher profit with its standard deviation decreased by 6 % as indicated in Table 2.

Table 1. Some price scenarios and the corresponding optimal production plans

No. of scenarios	Price change ($/kg)				Optimal production (kg)				Profit ($/10 days)
	Ethanol	Lactic acid	Itaconic acid	Citric acid	Ethanol	Lactic acid	Itaconic acid	Citric acid	
m (σ)*	0.728 (0.253)	1.639 (0.014)	2.270 (0.475)	1.459 (0.01)	21,000	10,900	9,600	0	44,289
1	-0.435	-0.025	0.389	0.015	6,300	8,720	11,000	1,920	37,875
2	0.314	-0.003	0.351	-0.006	22,000	8,720	11,000	1,920	55,271
3	-0.154	0.005	0.039	-0.004	12,600	10,900	7,680	3,840	38,949
4	-0.438	0	0.429	0.013	12,600	8,720	9,600	3,840	39,899
⋮									
1000	0.523	0.014	0.307	0.005	22,000	6,540	11,000	3,840	56,203

*m: mean; σ: standard deviation;
　data sources: http://www.the-innovation-group.com/ChemProfiles/,
　　　　　http://mindbranch.com/listing/product/R154-858.html

Table 2. Profit change by flexible operational planning

Profit	Fixed operation	Flexible operation
Mean, m ($/year)	1,623,826	1,630,743
Standard deviation, σ	7,557	7,104 (-6%)

The amount of required raw materials as well as the production ratios of each product k is determined according to the price scenarios. The prices of raw materials are fixed in this case to evaluate the effectiveness of the proposed operational planning. Subsequently, the profit variability is further reduced using the futures contracts of raw materials. The size of futures contracts demanded is determined to minimize the profit variance or to maximize the risk-return ratio for the price scenarios of raw materials and futures contracts. Figure 1 shows that the resulting cumulative profit curve after hedging

exhibits the steeper slope, i.e., the less variance of the profit, than the one before. Table 3 indicates that the variance reduction is 25.6 % of its standard deviation where the optimal hedge ratio is suggested to be 0.24. In this case, it turns out that the σ minimizing hedge ratio also maximizes the risk-return ratio (m/σ).

Table 3. Profit change by purchasing the futures of raw materials

Profit	Without hedging	With hedging	
		Minimizing σ	Maximizing m/σ
Mean, m ($/year)	1,937,735	1,942,920	1,942,920
Standard deviation, σ	2,027	1,508 (-25.6%)	1,508 (-25.6%)
Hedge ratio*	0	0.24	0.24

*hedge ratio = (size of the futures contracts) / (size of purchasing raw materials)
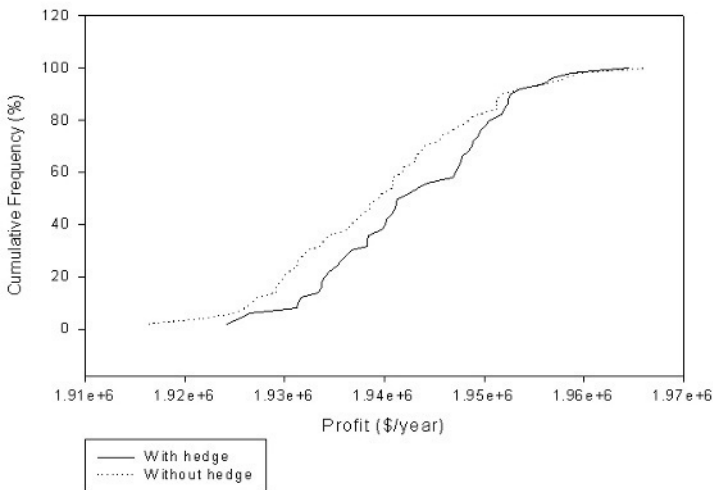
Figure 1. The cumulative profits before and after hedging using futures.

7. Conclusion

An integrated biorefinery process has been proposed to enhance the economic profitability of manufacturing and supplying the environmentally friendly products. The high profit variability of a biorefinery owing to the price fluctuations in the market could be reduced by diversifying the raw materials and products as well as by purchasing the futures contracts. The operational and financial planning model for the integrated biorefinery process has been described with an illustrative example. A biorefinery could make higher profit from the flexible production of ethanol and biochemicals. Furthermore, the profit variation was evaluated from historical data so that the procurement planning on the raw materials and futures could be established to prepare for the expected risks. The operation of the proposed integrated planning process will contribute to escalating the efficiency and flexibility of biorefineries in volatile environments.

References

G. Franceschin, Zamboni, A., Bezzo, F. and Bertucco, A., 2008, Ethanol from corn: a technical and economical assessment based on different scenarios. Chem Eng Res Des 86: 488-498.

J.C. Hull, 2005, Fundamentals of futures and options markets, 5th edition. (Prentice Hall, NJ, USA).

H.G. Joglekar, Rahman, I., Babu, S., Kulkarni, B.D. and Joshi, A., 2006, Comparative assessment of downstream processing options for lactic acid. Sep Purif Technol, 52: 1-17.

B. Kamm, and Kamm, M., 2007, Biorefineries – multi product processes. Adv Biochem Engin/Biotechnol, 105: 175-204.

D.E. Kenyon, 2001, Virginia basis tables for corn, soybeans, and wheat. Agricultural and Applied Economics, Publication Number 448-016.

A.A. Koutinas, Arifeen, N., Wang, R. and Webb, C., 2007, Cereal-based biorefinery development: integrated enzyme production for cereal flour hydrolysis. Biotechnol Bioeng, 97: 61-72.

A.A. Koutinas, Wang, R. and Webb, C., 2004, Restructuring upstream bioprocessing: technological and economical aspects for production of a generic microbial feedstock from wheat. Biotechnol Bioeng, 85: 524-538.

J.R. Kwiatkowski, McAloon, A.J., Taylor, F. and Johnston, D.B., 2006, Modeling the process and costs of fuel ethanol production by the corn dry-grind process. Ind Crops Prod, 23: 288-296.

K. Lakhdar, and Papageorgiou, 2008, An iterative mixed integer optimization approach for medium term planning of biopharmaceutical manufacture under uncertainty. Chem Eng Res Des, 86: 259-267.

Z. Li, Ding, S., Li, Z. and Tan, T., 2006, L-Lactic acid production by *Lactobacillus casei* fermentation with corn steep liquor-supplemented acid-hydrolysate of soybean meal. Biotechnol J, 1: 1453-1458.

A.C. Lim, Zhou, Y., Washbrook, J., Sinclair, A., Fish, B. et al., 2005, Application of a decision-support tool to assess pooling strategies in perfusion culture processes under uncertainty. Biotechnol Prog 21: 1231-1242.

National Agricultural Statistics Service (NASS), Agricultural Statistics Board, U.S. Department of Agriculture, (July 2008). Agricultural prices.

J. Park, S. Park, C. Yun and Y. Kim, 2009, An integrated model for financial risk management and optimal operation of a refinery, Ind Eng Chem Res, submitted.

D. Porro, Bianchi, M.M., Brambilla, L., Menghini, R., Bolzani, D. et al., 1999, Replacement of a metabolic pathway for large-scale production of lactic acid from engineered yeasts. App Env Microbiol, 65: 4211-4215.

J. Sadhukhan, Mustafa, M.A., Misailidis, N., Mateos-Salvador, F., Du, C. and Campbell, G.M., 2008, Value analysis tool for feasibility studies of biorefineries integrated with value added production. Chem Eng Sci, 63: 503-519.

N.E. Sammons Jr., Yuan, W., Eden, M.R., Aksoy, B. and Cullinan, H.T., 2008, Optimal biorefinery product allocation by combining process and economic modeling, Chem Eng Res Des, 86: 800-808.

E.S. Schwartz, 1997, The stochastic behavior of commodity prices: implications for valuation and hedging. J Finance, 52: 923-973.

M.A. Steffens, Fraga, E.S. and Bogle, I.D.L., 2000, Synthesis of bioprocesses using physical properties data. Biotechnol Bioeng, 68: 218-230.

T. Tran, Patel, T., Iland, T. and Truong, J., 2004, OU biorefining technical report for biomass production.

C. Vishnu, Seenayya, G. and Reddy, G., 2000, Direct conversion of starch to L(+) lactic acid by amylase producing *Lactobacillus amylophilus* GV6. Bioproc Eng, 23: 155-158.

Th. Willke, and Vorlop, K.-D., 2001, Biotechnological production of itaconic acid. Appl Microbiol Biotechnol, 56: 289-295.

C. Yun, Y. Kim, J. Park and S. Park, 2009, Optimal procurement and operational planning for risk management of an integrated biorefinery process, Chem Eng Res Des, accepted.

A Dynamical Model for the Fermentative Production of Fructooligosaccharides

O. Rocha[a,b], C. Nobre[a], A. Dominguez[a], D. Torres[a,b], N. Faria[b], L. Rodrigues[a,b], J.A. Teixeira[a], E.C. Ferreira[a] and I. Rocha[a,b]

[a]*IBB - Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, PORTUGAL*
[b]*Biotempo, Lda., Avepark – Zona Industrial da Gandra, Apartado 4152, 4806 – 909, Caldas das Taipas, Braga, Portugal*

Abstract

In this paper a detailed mathematical model is presented for the fermentative production of fructo-oligosaccharides with *Aspergillus* sp. The model accounts for hydrolysis and transfructolization reactions, as well as biomass formation and it contains 27 parameters that were determined from experimental data using a System Biology toolbox with the Simulated Annealing method for curve fitting. Several additional experiments were performed in bioreactors where the time variation of 7 state variables (Sucrose, Glucose, Fructose, 1-Kestose, Nystose, 1-fructosyl nystose and Biomass) was measured. Experimental data were compared with results from simulations using the estimated parameters and it was verified that the model can predict the FOS production profile. The good agreement between simulated and experimental data was verified by calculating the relative percentage deviation modulus, which was lower than 10% for all cases except one. The derived and validated model can be used for process optimization, for example for indicating which fed-batch strategy could be used to improve the production of FOS while minimizing glucose concentration.

Keywords: Modelling, Simulation, Fructooligosaccharides.

1. Introduction

Within Industrial Biotechnology, a very promising application is the production of ingredients for functional foods, since the market for those products has been growing at very interesting rates [9]. In recent years some prebiotics have been described as beneficial food ingredients because of their proprieties of modifying the intestinal microbiota, favoring the growth of some beneficial bacteria [1;2;10;11]. Fructo-oligosaccharides (FOS) have become one of the most important prebiotic products with healthy properties, being possible to find them, usually in trace amounts, as natural components in fruits, vegetables and honey [8;11]. Although industrially these products are mainly extracted from those natural sources, they can be also produced from sucrose by the action of β- fructofuranosidase [FFase; EC 3.2.1.261] obtained from some organisms. Various fungi such as *Aureobasidum* sp., *Aureobasidum pullulans*, *Aspergillus niger*, *Aspergillus japonicas*, *Aspergillus oryzae* and *Scopulariopsis brevicaulis* [3;5;6;8;12;13;15] produce those oligosaccharides that are mainly composed of 1-Kestose, Nystose, and Fructosylfuranosyl nystose in which 1-3 fructose units are bound at the β-2,1 position of sucrose [16].

However, besides the fructosyltransferase activity, β-fructofuranosidase also exhibits hydrolytic activity [4;5;7;12], which can dominate the process depending on a combination of factors including the sucrose concentration. This fact will ultimately lead to lower production yields and to a contamination of the final product with the monosaccharides glucose and fructose. Additionally, in a fermentative process for the production of FOS, substrate consumption for biomass growth has also to be considered, increasing even further the complexity of the process and motivating the application of mathematical modelling approaches such that non-obvious operation conditions can subsequently be found that maximize the productivity of FOS and minimize the accumulation of monosaccharides.

2. Mathematical Model

The main aim of this work was to formulate a general model that characterizes the main reactions representing the fermentative FOS production process. It is based on the empirical equations of enzymatic production of fructooligosaccharides from sucrose. The model contemplates both hydrolysis and transfructosylation kinetic equations representing β- fructofuranosidase activity and growth rate equations for the microorganism. The enzymatic reactions were divided in two main categories: the hydrolysis reactions, representing FOS and sucrose degradation, and the transfructosylation reactions that describe FOS synthesis. In the formulation of the model only three different FOS were considered: 1-Kestose, Nystose, 1- fructofuranosyl nystose. A prior analysis of fermentation samples by HPLC indicated only the presence of these oligosaccharides.

2.1. Hydrolysis reactions

The hydrolysis of saccharose and FOS by β- fructofuranosidase is described by equations 2 to 5. It is considered that all the di- and oligo-saccharides can be hydrolysed by the action of the enzyme.

$$GF \xrightarrow{r1} k_1.G + k_2.F \tag{1}$$

$$GF_2 \xrightarrow{r2} k_3.GF + k_4.F \tag{2}$$

$$GF_3 \xrightarrow{r3} k_5.GF_2 + k_6.F \tag{3}$$

$$GF_4 \xrightarrow{r4} k_7.GF_3 + k_8.F \tag{4}$$
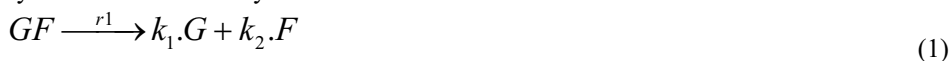
Duan and co-authors [4] proposed a Michaelis-Menten equation with substrate inhibition to represent nystose hydrolysis. In our model, this phenomenon was considered to occur also during the hydrolysis of 1-kestose and 1-Fructosylfuranosyl nystose. The FOS hydrolysis kinetic equation is then given by:

$$r_i = \frac{Vmh_{GFi} \times GF_i}{GF_i(1 + \dfrac{GF_i}{Kih_{GFi}}) + Kmh_{GFi}} \tag{5}$$

with i = 2, 3, 4, where r_i is the ith fructooligosaccharide hydrolysis rate (g L^{-1} h^{-1}), Vmh_{GFi} is the maximum hydrolysis rate (g L^{-1} h^{-1}), GF_i is the concentration (g L^{-1}) of nystose, 1-kestose or 1-Fructosylfuranosyl nystose, Kih_{GFi} is the substrate inhibition constant (g L^{-1}), and Kmh_{GFi} is the Michaelis-Menten constant (g L^{-1}) for GF_i.

For sucrose hydrolysis, a Michaelis-Menten equation was used, given by:

$$r_1 = \frac{Vmh_{GF} \times GF}{Kmh_{GF} + GF}$$

(6)

where Vmh_{GF} is the maximum hydrolysis rate (g L^{-1} h^{-1}), GF is the sucrose concentration (g L^{-1}) and Kmh_{GF} is the Michaelis-Menten constant for sucrose (g L^{-1}).

2.2. Transfructosylation reactions

The formation of oligosaccharides can occur by a transition of a fructosyl residue from one molecule to another like it was described by Duan and co-authors [4]:

$$2GF \xrightarrow{\ r5\ } k_9.GF_2 + k_{10}.G$$

(7)

$$2GF_2 \xrightarrow{\ r6\ } k_{11}.GF_3 + k_{12}.GF$$

(8)

$$2GF_3 \xrightarrow{\ r7\ } k_{13}.GF_4 + k_{14}.GF_2$$

(9)

In sucrose transfructosylation, represented by equation 7, a Michaelis-Menten equation with substrate inhibition and competitive glucose inhibition was used:

$$r_5 = \frac{VmT_{GF} \times GF}{GF(1+\dfrac{GF}{Ksts}) + Kmst(1+\dfrac{G}{Kgst})}$$

(10)

where r_5 is the sucrose transfructosylation rate (g L^{-1} h^{-1}), VmT_{GF} is the maximum transfructosylation rate (g L^{-1} h^{-1}), GF is the sucrose concentration (g L^{-1}), $Ksts$ is the substrate inhibition constant (g L^{-1}) for sucrose as a substrate, $Kgst$ is the competitive inhibition constant (g L^{-1}) for glucose and $Kmst$ is the Michaelis-Menten constant (g L^{-1}) for sucrose.

Equations 8 and 9 represent the Nystose and 1-kestose transfructosylation reactions. A competitive glucose inhibition term was also included in the Michaelis-Menten equation, since Duan [4] reported this phenomenon for these two fructooligosaccharides. Equation 11 represents the fructooligosaccharides transfructosylation reaction rates:

$$r_j = \frac{VmT_{GFi} * GF_i}{GF_i + Kmt_{GFi}(1+\dfrac{G}{Kit_{GFi}})}$$

(11)

with i =2, 3; j = 6, 7; where VmT_{GFi} is the maximum transfructosylating rate (g L^{-1} h^{-1}), GF_i is the FOS concentration (g L^{-1}), Kmt_{GFi} is the Michaelis-Menten constant (g L^{-1}) for the GF_i oligosaccharide and Kit_{GFi} is the competitive inhibition constant (g L^{-1}).

2.3. Growth reactions

The formation of biomass can either occur from glucose or fructose consumption and can be described by:

$$Y_G G \xrightarrow{\ r_8\ } X$$

(12)

$$Y_F F \xrightarrow{\ r9\ } X$$

(13)

Since substrate consumption for maintenance was considered to be significantly smaller, it was neglected. The proposed Monod equations are given as follows:

$$r_j = \frac{\mu_{j,max}.S_j.X}{S_j + KS_j}$$

(14)

r_j = 8, 9; where r_j is the growth rate of the microorganism (g L^{-1} h^{-1}), $\mu_{j,max}$ is the maximum specific growth rate (h^{-1}) on glucose or fructose, S_j is the glucose or fructose concentration (g L^{-1}), X is the biomass concentration (g L^{-1}) and KS_j is the affinity constant for the substrate (g L^{-1}). The Y_G and Y_F in equations 12 and 13 are the biomass yields when the glucose or fructose are used for biomass growth (g g^{-1}).

2.4. Derivation of model equations

After establishing both the reaction scheme and the kinetic equations, a general dynamical model of the process accounting for mass transfer, biomass growth and enzymatic reactions was defined. In the formulation of this model 7 state variables have been considered: sucrose (GF), Glucose (G), Fructose (F), 1-Kestose (GF_2), Nystose (GF_3), 1-fructofuranosyl nystose (GF_4) and Biomass (X).

The time derivatives of the concentration of studied components for a fed-batch bioreactor are given as:

$$\frac{dGF}{dt} = (-r_1 + k_3 \cdot r_2 - r_5 + \frac{k_{12}}{2} \cdot r_6) + \frac{F_{in}}{V}GF_{in} - D \cdot GF \tag{15}$$

$$\frac{dG}{dt} = (k_1.r_1 + \frac{k_{10}}{2}.r_5 - Y_G.r_8) - D.G \tag{16}$$

$$\frac{dF}{dt} = (k_2.r_1 + k_4.r_2 + k_6.r_3 + k_8.r_4 + Y_F.r_9) - D.F \tag{17}$$

$$\frac{dGF_2}{dt} = (-r_2 + k_5.r_3 + \frac{k_9}{2}.r_5 - r_6 + \frac{k_{14}}{2}.r_7) - D.GF_2 \tag{18}$$

$$\frac{dGF_3}{dt} = (-r_3 + k_7.r_4 + \frac{k_{11}}{2}.r_6 - r_7) - D.GF_3 \tag{19}$$

$$\frac{dGF_4}{dt} = (-r_4 + \frac{k_{13}}{2}.r_7) - D.GF_4 \tag{20}$$

$$\frac{dX}{dt} = (r_8 + r_9) - D.X \tag{21}$$

where $F_{in,s}$ is the volumetric flow rate of sucrose feeding solution (L h^{-1}); GF_{in} is the sucrose concentration on the feeding (g L); D is the quotient between the total feed rate ($F_{in,total}$) and the V is the total volume of liquid inside reactor (L).

3. Materials and Methods

Two fermentations of *Aspergillus* sp. were performed in a 5 L fermentor (B. Braun Biotech International), model Micro-DCU 200 at a pH of 5 and 30ºC, with Czapek Dox Media of OXOID and an initial sucrose concentration of 200 grams per litre. During those fermentations, the time evolutions of the 7 state variables present in model were measured. The growth was monitored by dry cell weight, where three 10 mL samples were filtered with a 0.45 micron filter and dried at 105 ºC for 20 h. The supernatant was used to determine the carbohydrates concentration. They were analyzed in a JASCO HPLC instrument with a refractive index detector using a VARIAN MetaCarb 87P column. The column was maintained at 25 ºC, and a mixture of water and acetonitrile was used as a mobile phase at 1 mL min^{-1}.

Afterwards, the collected data were used for the determination of unknown kinetic and yield coefficients with a Simulated Annealing method included in the System biology toolbox [14]. For the estimation of kinetic and yield coefficients only data from one experiment of each microorganism were used. This experiment was chosen randomly. Data from the remaining experiments were used to validate the model accuracy.

The fitting was performed by minimizing a total cost function that represents the adjustment between experimental and simulated data:

$$Total \cos t = \sum_{i=1}^{n} \left(\frac{1}{N_p} \sum_{j=1}^{p} \left(\frac{\xi_{sim,ij} - \xi_{\exp,ij}}{\overline{\xi}_{\exp,ij}} \right)^2 \right)$$

(1)

where $\xi_{sim,ij}$ represents the simulated data and $\xi_{\exp,ij}$ is the experimental data for every point (p) for a given state variable (n) and Np is the total number of data points. The difference is divided by an average value $\overline{\xi}_{\exp,ij}$ with the purpose of attributing the same importance to all state variables.

4. Results and Discussion

The first task was to find the values for the 27 unknown parameters already described in model equations. The parameters obtained using the system biology toolbox were compared with some kinetic parameters values from literature [4], for a β-fructofuranosidase derived from an *Aspergillus japonicus*. Subsequently a simulation was carried out in System Biology toolbox using those parameters and the results were compared with the experimental data obtained from a second fermentation. The comparison between simulated and experimental data is shown in figure 1. These results show that the proposed mathematical model can predict correctly the time profiles for the state variables.

5. Conclusions

In this paper we present a detailed mathematical model for the production of fructo-oligosaccharides with *Aspergillus* sp. To the best of our knowledge, although several models representing the enzymatic reactions have been published, this is the first model that represents the fermentative process, therefore accounting for biomass formation.

Experimental data were compared with results from simulations using the estimated parameters and it was verified that the models can predict the FOS production.

Once the model is derived and validated, it can now be used for process optimization, for example for indicating which fed-batch strategy could be used to improve the production of FOS while minimizing glucose concentration.

References

1 Angeliki Kapiki, Christos Costalos, Christina Oikonomidou, Antigoni Triantafyllidou, Erini Loukatou and Vassiliki Pertrohilou. (2006) The effect of a fructo-oligosaccharide supplemented formula on gut flora of preterm infants. Early Human Development.

2 Bielecka,M., Biedrzycka,E., Majkowska,A., Juskiewicz,J. and Wroblewska,M. (2002) Effect of non-digestible oligosaccharides on gut microecosystem in rats. Food Research International, 35, 139-144.

3 Cruz,R., Cruz,V.D., Belini,M.Z., Belote,J.G. and Vieira,C.R. (1998) Production of fructooligosaccharides by the mycelia of Aspergillus japonicus immobilized in calcium alginate. Bioresource Technology, 65, 139-143.

4 Duan,K.J., Chen,J.S. and Sheu,D.C. (1994) Kinetic studies and mathematical model for enzymatic production of fructooligosaccharides from sucrose. Enzyme and Microbial Technology, 16, 334-339.

5 Fernandez,R.C., Maresma,B.G., Juarez,A. and Martinez,J. (2004) Production of fructooligosaccharides by beta-fructofuranosidase from Aspergillus sp 27H. Journal of Chemical Technology and Biotechnology, 79, 268-272.

6 Hatakeyama,Y., Takeda,H., Ooi,T. and Kinoshita,S. (1996) Kinetic parameters of beta-fructofuranosidase from Scopulariopsis brevicaulis. Journal of Fermentation and Bioengineering, 81, 518-523.

7 Kim,M.H., In,M.J., Cha,H.J. and Yoo,Y.J. (1996) An empirical rate equation for the fructooligosaccharide-producing reaction catalyzed by beta-fructofuranosidase. Journal of Fermentation and Bioengineering, 82, 458-463.

8 Lee,W.C., Chiang,C.J. and Tsai,P.Y. (1999) Kinetic modeling of fructo-oligosaccharide production catalyzed by immobilized beta-fructofuranosidase. Industrial & Engineering Chemistry Research, 38, 2564-2570.

9 Menrad,K. (2003) Market and marketing of functional food in Europe. Journal of Food Engineering, 56, 181-188.

10 Michael,B. (2002) Relationship of prebiotics and food to intestinal microflora. European Journal of Nutrition, V41.

11 Mikkelsen,L.L., Jakobsen,M. and Jensen,B.B. (2003) Effects of dietary oligosaccharides on microbial diversity and fructo-oligosaccharide degrading bacteria in faeces of piglets post-weaning. Animal Feed Science and Technology, 109, 133-150.

12 Sangeetha,P.T., Ramesh,M.N. and Prapulla,S.G. (2004) Production of fructo-oligosaccharides by fructosyl transferase from Aspergillus oryzae CFR 202 and Aureobasidium pullulans CFR 77. Process Biochemistry, 39, 755-760.

13 Sangeetha,P.T., Ramesh,M.N. and Prapulla,S.G. (2005) Recent trends in the microbial production, analysis and application of Fructooligosaccharides. Trends in Food Science & Technology, 16, 442-457.

14 Schmidt,H. and Jirstrand,M. (2006) Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology. Bioinformatics, 22, 514-515.

15 Sheu,D.C., Duan,K.J., Cheng,C.Y., Bi,J.L. and Chen,J.Y. (2002) Continuous production of high-content fructooligosaccharides by a complex cell system. Biotechnology Progress, 18, 1282-1286.

16 Shin,H.T., Baig,S.Y., Lee,S.W., Suh,D.S., Kwon,S.T., Lim,Y.B. and Lee,J.H. (2004) Production of fructo-oligosaccharides from molasses by Aureobasidium pullulans cells. Bioresource Technology, 93, 59-62.
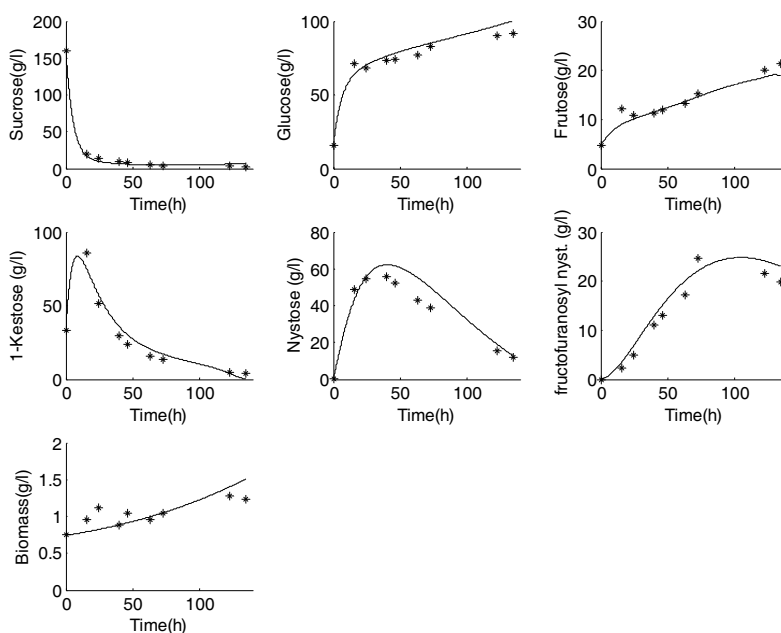
Figure 1: Comparison of *Aspergillus sp.* simulated and the experimental data from Ferm1. The simulation was performed with the parameters calculated from Ferm2.

Energy Consumption Minimization in Bioethanol Dehydration with Supercritical Fluids

Cecilia I. Paulo, [a] M. Soledad Diaz, [a] Esteban A. Brignole[a]

[a] *Planta Piloto de Ingeniería Química (PLAPIQUI), Universidad Nacional del Sur-CONICET. Camino de La Carrindanga Km 7, Bahía Blanca 8000, Argentina.*

Abstract

In this work, we propose a rigorous model for bioethanol dehydration process with supercritical propane to minimize energy consumption. Thermodynamic predictions are performed with an upgraded Group Contribution with Association Equation of State, GCA-EOS. As compared to the basic scheme for dehydration with supercritical fluids, vapor recompression, as well as feed preconcentration could be highly energy efficient. We further consider alternative integration schemes between process streams, associated to different nonlinear programming problems. Special attention has been devoted to a new scheme that integrates the vapor recompression scheme to the preconcentration step, which provides additional reduction in total energy consumption. We demonstrate that bioethanol dehydration can be a sustainable alternative that is energetically competitive with molecular sieves in the production of this biofuel.

Keywords: Bioethanol, dehyextraction, near critical fluid, NLP.

1. Introduction

Alternatives to fossil fuels are being investigated to reduce the world's dependence on non-renewable resources. Biofuels are currently considered as relevant sustainable technologies due to energy security reasons, environmental concerns, foreign exchange savings, and socioeconomic issues related to the rural sector. The reduction of greenhouse gases pollution is the main advantage of utilizing biomass energy. The most common renewable fuel is ethanol derived from corn grain (starch) and sugar cane (sucrose). Wood, straw and even household wastes may also be economically converted to bioethanol. However, there is need for decreasing energy consumption in the entire bioethanol supply chain to make it economically competitive with fossil fuels. Much research is being pursued on the use of lignocellulosic biomass as an attractive feedstock for future supplies of ethanol. On the other hand, downstream processes of bioethanol separation and dehydration are being studied. Karuppiah et al. (2008) have proposed different design alternatives for the transformation of corn kernels to fuel ethanol, using distillation together with molecular sieves and adsorption units with corn grits to achieve fuel-grade bioethanol. The use of pervaporation membranes has been also analyzed as an alternative to extractive distillation and molecular sieves (Hoch & Espinosa, 2008). The use of light hydrocarbons as supercritical solvents for bioethanol dehydration has been proposed as a low energy consumption technology (Brignole et al., 1987, Horizoe et al., 1993). The basic process consists of two steps, the extraction of bioethanol from the aqueous solution with a near critical solvent, and a final separation of ethanol from the solvent in a distillation train. The light hydrocarbon solvent has good selectivity for ethanol and water-solvent relative volatility becomes greater than one at the solvent recovery column (water entrainment effect). The inclusion of

different alternatives to the basic scheme has been formulated as a mixed integer nonlinear programming (MINLP) problem whose solution has provided improved energy consumption options (Gros et al., 1998, Diaz et al., 2000).

In this work, we further consider alternative integration schemes between process streams within different nonlinear programming (NLP) problems. In particular, we analyze a new scheme that integrates the vapor recompression scheme to the pre-concentration step, which can provide additional reduction in total energy consumption. Thermodynamic predictions are performed with an upgraded Group Contribution with Association Equation of State, GCA-EOS (Jorgensen, 1988; Gros et al., 1996, Ferreira et al, 2004). Numerical results show that process economics are comparable to the use molecular sieves.

2. Process Description

2.1. Basic dehydration with supercritical fluids

In a basic bioethanol dehydration process with supercritical light hydrocarbons (Diaz et al., 2000), the main units are the high pressure extractor and solvent recovery columns, as shown in Fig. 1. The supercritical fluid solvent stream enters the extraction column at the bottom, while the ethanol-water mixture is fed to the top of the column. The column is operated at conditions near the critical temperature of the solvent and at pressure above the critical. The extract (mainly ethanol and solvent) contains a small amount of water. The raffinate mainly consists of water and a very small amount of bioethanol and almost no solvent. The extract is reduced in pressure through a valve and fed to a distillation column to recover the solvent. In this column, the solvent is recovered and the complete dehydration of bioethanol is obtained, through entrainment of water by the near critical solvent. The distillate is returned as the supercritical solvent to the extractor and almost absolute bioethanol is obtained as bottoms product in the distillation unit. In this process, numerical results strongly depend on the accuracy of the thermodynamic model predictions of key phase equilibrium properties.
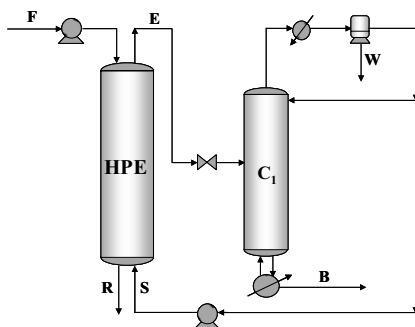
Figure 1. Basic extraction-dehydration scheme. HPE: high-pressure extractor, C_1: dehydration column, F: aqueous feed, R: raffinate, B: dehydrated bioethanol, S: solvent.

2.2. Alternative process schemes

Figure 2 shows the extraction-dehydration process superstructure that embeds process schemes analyzed in this work. The recovery of ethanol from dilute aqueous solutions can be achieved with low energy consumption by ordinary distillation, if the separation goal is the complete alcohol removal from the solution (and not dehydrated ethanol). In this case, the high alcohol-water relative volatility for dilute aqueous mixtures, allows

the increase of ethanol concentration using a simple stripping column. Therefore, *preconcentration* (PC) of the process feed reduces the flowrate of the aqueous solution to the extractor and, consequently, supercritical solvent requirements. The use of a two solvent recovery columns scheme (C1+C2) makes possible energy integration between the feed preconcentration and first solvent recovery column, such as matching the top vapor from the preconcentrator (C) to the reboiler of the first solvent recovery column (HE1). However, the stream matches depend on the preconcentrator operating pressure. There is an alternative vapor *recompression* scheme (RC), in which energy consumption is mainly determined by the compression work provided to the overhead vapor of the first distillation column (C1). The required energy in the column reboiler/condenser (HE3) is supplied by the condensation of the recompressed vapor (V_{C1}). An alternative scheme analyzed in this work is to use a turbine (TC) as driver for the compressor (RC) and integrate the exhaust steam stream to the preconcentrator reboiler (HE1). The turbine operates with middle pressure steam. No external heating services are thus required. The use of vapor recompression is justified from the low temperature difference between top and bottom at this column, achieved when a rather high concentration of solvent is kept in the bottom product. This gives low energy consumption for vapor recompression at the expense of some additional energy consumption by conventional heating in the second separating column.
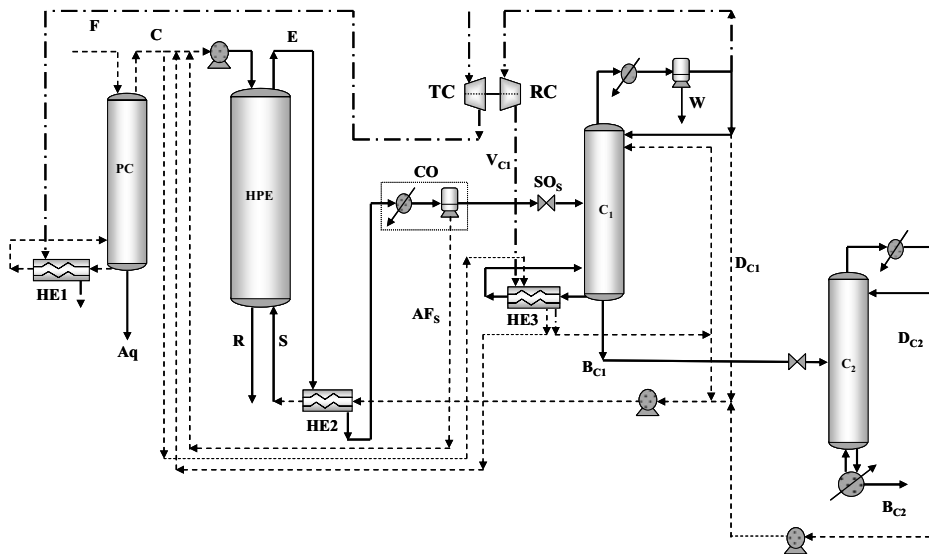
Figure 2. Superstructure for extraction-dehydration process. HPE: high-pressure extractor, C1: dehydration column, C2: second solvent recovery column, F: aqueous feed, R: raffinate, B: dehydrated bioethanol, S: solvent, CO: cooling unit, SOs: solvent organic stream, AFs: aqueous feedback stream, PC: preconcentrator, RC: recompressor, HE1: preconcentrator reboiler; HE2: solvent heater, HE3: column C1 reboiler. Full line: basic flowsheet; dashed line: alternative schemes; dashed point line: optimal scheme.

3. Mathematical model for extraction-dehydration with supercritical fluids

Different process alternatives have been formulated as a series of NLP problems. Main design variables are extraction temperature (TE) and pressure (PE), solvent flow rate (S) and reflux ratio at dehydration first column (RC1). Their bounds are shown in Table 1.

Table 1. Design variable bounds

Variable	Lower bound	Upper bound
Extractor Temperature, T_E (K)	325.00	420.00
Extractor Pressure, P_E (bar)	40.00	100.00
Reflux ratio, R_{C1}	0.30	2.50
Solvent, S (kmol/h)	45.00	1500.00

The process mathematical model (Diaz et al., 2000) includes first principles rigorous models for high-pressure multistage extractors (Kehat & Ghitis, 1981), low and high-pressure distillation columns (Naphtali & Sandholm, 1971) and a multiphase flash (Michelsen, 1982). The key thermodynamic properties of the extraction-dehydration process are based on the Group Contribution Equation of State with Association, GCA-EOS model (Skjold-Jorgensen, 1988, Gros et al, 1996) that provides reliable phase equilibrium predictions at high pressure in mixtures with association. Operating bounds and process specifications have been included as inequality constraints, as shown in Table 2.

Table 2. Inequality constraints

Unit	Constraint	Description	Bound
Extractor	r1	Ethanol recovery (%)	≥ 98.50
Solvent recovery column C_1	r2	Water composition in top vapor phase of C_1	$\leq Y_{H2O(sat)}$
Solvent recovery column C_1	r3	Ethanol recovery (%)	≥ 98.00
Solvent recovery column C_1	r4	Ethanol (solvent free basic) in bottom of C_1	≥ 99.00
Preconcentrator	r5	Ethanol recovery (% molar)	≥ 99.50
Preconcentrator	r6	Energy available from preconcentrator vapor	$\geq QreboilerC_1$
Preconcentrator	r7	Preconcentrator vapor-reboiler C_1 temperature difference	≥ 15.00

The goal of this work is to minimize the dehydration process energy consumption to provide an economically attractive clean technology. The objective function is composed of several terms corresponding to pumping energy for liquids (solvent and aqueous feed) and heating requirement in distillation columns (PC, C1, C2), as well as their integration in the different proposed flowsheet schemes. Mechanical energy (kJ/kg) has been affected by a factor of 3.0 (Streich & Bolkart, 1982) to convert it to an amount of thermal energy of equivalent cost, so as evaluate the different alternatives on a similar cost basic. Nonlinear programming problems have been solved with a Successive Quadratic Programming algorithm (Biegler and Cuthrell, 1985).

4. Discussion of results

In this work, we consider the dehydration process downstream the fermentation step in a bioethanol plant. A typical feed to this sector is 10,000 kg/h of an aqueous solution with

10 wt% ethanol concentration. Propane is the near critical solvent (Tc=369,8, Pc=41,9 bar). Units specifications are: *extractor,* 10 stages; *first solvent recovery column*, 35 stages, pressure: 25 bar; *second solvent recovery column*: pressure: 12 bar; reflux ratio, 0,70. We have formulated four nonlinear programming (NLP) problems corresponding to the most attractive process schemes determined in previous work (Diaz et al., 2000) and a new integration scheme between turbo-compressor and preconcentrator reboiler. Table 3 shows optimal conditions and minimum energy consumption for each scheme. As compared to the basic supercritical extraction-dehydration scheme, vapor recompression or feed preconcentration can be highly energy efficient (columns 2 and 3 in Table 3). The last column in Table 3 shows that the proposed scheme integrating the exhaust steam stream from the turbine to the preconcentrator reboiler gives the optimal structure for the high pressure bioethanol dehydration process. This option improves the integrated energy consumption in about: 29% respect to the preconcentration option, 65% respect to the recompression option and 88% respect to the basic scheme. The optimal scheme also shows lower operating temperature and pressure in the extractor, and an important reduction in solvent requirement, as compared to the remaining schemes.

Table 3 Optimal operating conditions and energy consumptions for each alternative process.

	Basic	Vapor Recompression	Preconcentration	Preconcentration + Recompression
T_E (K)	397	393.06	362.81	361.28
S (kmol h^{-1})	760	820.71	194.50	200.00
R_{C1}	1.50	1.50	0.98	0.90
P_E (bar)	100	78.93	55.00	55.00
Ethanol purity (%)	99.00	99.99	99.00	99.00
Ethanol recovery in C_2	98.00	98.00	98.87	98.65
Preconcentrator (kJ kg^{-1})	-	-	3184.98	3188.84
Column C_1 (kJ kg^{-1})	22200.00	22284.46	3735.46	3697.64
Column C_2 (kJ kg^{-1})	500.00	261.33	254.28	255.45
Pumping (kJ kg^{-1})	3500.00	2511.72	291.67	299.18
Recompression (kJ kg^{-1})	-	5.995.32	-	325.73
Integrated energy consumption (kJ kg^{-1})	26200.00	8768.36	4346.79	3082.93

5. Conclusions

We have formulated nonlinear programming problems based on first principles rigorous models with reliable thermodynamic predictions by a Group Contribution Equation of State with association (GCA-EOS), to analyze different designs for bioethanol dehydration plants, using propane as the supercritical or near critical solvent. The GCA-EOS model gives reliable properties predictions of highly nonideal azeotropic mixtures at low and high pressure conditions. Four process schemes have been optimized (basic supercritical extraction-dehydration, vapor recompression, feed preconcentration and preconcentration + vapor recompression). Numerical results show that the

preconcentrator+recompression scheme provides a significant reduction in energy consumption, making the high pressure process comparable to the use of the well-known dehydration technology with molecular sieves. We are currently including both capital and operating costs in our analysis to perform a detailed cost comparison. Nevertheless, optimization results indicate that the supercritical extraction-dehydration of bioethanol can be energy efficient, helping to make bioethanol a sustainable and economically competitive alternative to non-renewable fossil fuels.

References

L. T. Biegler, J. E. Cuthrell, 1985, Improved Infeasible Path Optimization for Sequential Modular Simulators II. The Optimization Algorithm, Computers and Chemical. Enginering 9, 257-265.

E. A. Brignole, P. Andersen, Aa. Fredenslund, 1987, Supercritical fluid extraction of alcohol frim water, Ind. Eng. Chem. Res. 26, 254.

O. Ferreira, E.A. Brignole, E.A. Macedo, 2004, Modelling of phase equilibria for associating mixtures using an equation of state, J. Chem. Thermodynamics 36, 1105-1117.

S. Diaz, H. Gros, E. A. Brignole, 2000, Thermodynamic Modeling, Synthesis and Optimization of Extraction-Dehydration Processes, Computers & Chemical Engineering, 24, 9, 2069-2080.

H. P. Gros, S. Bottini, E. A. Brignole, 1996, A group contribution equation of state for associating mixtures, Fluid Phase Equilibria, 116, 537.

H. P. Gros, M.S. Diaz, E. A. Brignole, 1998, Near Critical Separation of Aqueous Azeotropic Mixtures: Process Synthesis and Optimization, J. of Supercritical Fluids, 12, 69.

M. P. Hoch, J. Espinosa, 2008, Conceptual Design and Simulation Tools Applied to the Evolutionary Optimization of a Bioethanol Purification Plant, Ind.Eng.Chem.Res., 47, 7381–7389.

H. Horizoe, T. Tanimoto, I. Yamamoto, Y. Kano, 1993, Phase equilibrium study for the separation of ethanol-water solution using subcritical and supercritical hydrocarbon solvent extraction, Fluid Phase Equilibria, 84, 297.

R. Karuppiah, A. Peschel, I. E. Grossmann, M. Martín; W. Martinson; L. Zullo, 2008, Energy Optimization for the Design of Corn-Based Ethanol Plants, AIChE J., 54 (6), 1499–1525.

E. Kehat, B. Ghitis, 1981, Simulation of an extraction column, Computers & Chemical Engineering, 5, 171.

M. L. Michelsen, 1982, The isotermal flash problem: part II: phase split calculations, Fluis Fhase Equilibria, 9, 21.

L. M. Naphtali, D. P. Sandholm, 1971, Multicomponent separation calculations by linearization, American Institute of Chemical Engineering Journal, 17, 148.

S. Skjold-Jorgensen, 1988, Group contribution equation of state (GC-EOS): a predictive method for phase equilibrium computations over wide ranges of temperatures and pressures up to 30 Mpa., Ind. Eng. Chem. Res., 27, 110.

M. Streich, A. Bolkart, 1982, Heat pumps and ORCs can effectively compete in waste-heat utilization projects, Oil & Gas Journal, 80, 186.

Corrosion Control Document Database System in Refinery Industry

Junghwan Kim[a], Sang-Rok Park[b] and Il Moon[a]

[a]*Dept. of Chemical and Biomolecular Engineering, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul 120-749, Korea*
[b]*GS Catex Corporation, 1056, Wollae-dong, Yeosu-si, Jeollanam-do 555-260, Korea*

Abstract

This paper focuses on techniques of improving refinery reliability, availability and profitability. Our team developed a corrosion control document (CCD) database system for refinery process. This system includes a method for defining and measuring the component risk and it also provides a powerful tool for managing many of the important elements of a refinery plant. The simulation has been applied for this system to find the critical component without direct measurement and experiment. It is possible to expect the risk parts of the refinery process.

Recent study shows the loss due to corrosion in US is around $276 billion. It's a big concern for both managers and engineers of refinery industry. Therefore it is essential to make the profitable and convenient system for refinery industry to improve reliability.

CCD database system consists of numerous parts namely damage mechanism (DM), design data, critical reliability variable (CRV), guidelines, etc. CCD database system has been developed on the basis of the corrosion control in refinery industry. It also improves the safety of refinery process and reduces the cost of corrosion greatly

Keywords: Corrosion Control Document (CCD), refinery process, damage mechanisms (DM), critical reliability variables (CRV)

1. Introduction

As an oil refinery has become bigger and accumulated recently, an accident of corrosion are more likely to be extended to serious accident. In fact, refinery process accidents have been happened continuously in the inside and outside of country. These accidents bring about many casualties and economic loss. Recently, an American report shows that corrosion cost was around $276 billion in 1998, and it was large sum accorded to 3.1 percent of American GDP of that year.

The principal case of corrosion accident in the refinery process was storage tank explosion in June 17, 2001 at Delaware. The storage tank exploded due to high-pressure gas release caused by the corrosion of connection pipe in tank. In consequence, one person who belong to repair works contractor died and eight people were wounded in this accident.

In case of Korea, the accident which took place in Oct 20, 2003 was caused by high temperature sulfidation in HOU unicracking process. Corrosion made the tube thin. With thinned tube break caused by internal pressure, distilled crude oil in tube flew out, and then the fire broke out due to heater's high fever.

Like this, corrosion accidents occur frequently in the world, and corresponding men and money loss has been serious. Therefore, systematic system is required so as to reduce

accidents caused by corrosion and to minimize corrosion cost. First of all, CCD (corrosion control document) development is necessary for development of systematic system. However, while most of internal oil companies state a few reasons such as lack of work force and time, they postpone the development of CCD.

In this study, our team developed CCD in oil refinery process and systemized development process. As following these, oil refining process as well as corrosion condition will be improved. Furthermore, on the basis of data from CCD DB, if this and currently used RBI (risk based inspection) apply to actual case together, economic effect can be achieved.

2. CCD Development

2.1. Background

In many cases, test and check of corrosion accident, in recent refined oil service as well as in petroleum chemistry equipment, was approximately recorded or it was not recorded at all. Also, information about process and arrangement which are apt to be corroded has been not well managed. Even though loss by corrosion is very small or negligible, it should be described exactly or managed well. As well, several things have to be determined, that is, which part is protected from record of loss caused by corrosion and which thing is prevented. Therefore, CCD development in refined oil and petrochemical process is needed firstly.

In this study, we developed CCD with an emphasis on CDU process in an oil refinery. Generally, Corrosion Control Document includes the following contents.

- *Inclusive document about equipment efficiency decrease*
- *Unit process & Process State*
- *Shutdown & startup condition that is related to steady-state & corrosion*
- *Process Flow Diagram, materials, corrosion diagram*
- *Corrosion control loops (Total system, Slurry system, Reflux System)*
- *Potential case of efficiency decrease in unit process and the spot where the accident took place*
- *Efficiency decrease mechanism by corrosion*
- *Corrosion control procedures, injections, inhibitors inspection, corrosion monitoring, process alteration and advice of material replacement*
- *Integrity Operation Windows (IOW)*

CCD is a document that includes information about from principle design of equipment to corrosion reason, corrosion mechanism, past record, guide line on corrosion prevention. Hence, CCD development is requisite for equipments in an oil refinery as well as petrochemical plant. Also, it is fundamental document to improve reliability of equipment and to reduce loss by corrosion.

2.2. Process

The general order of CCD development is constituted of 4 steps like following Fig.1.

The first step is to classify the quality of material according to equipment in PFD process. After that, describe DM (damage mechanism) which gives effect on equipment with each process. On the basis of that, select CRV (critical reliability variables) which

influence on inspection interval and equipment integrity. The last step is to make operation guidelines which propose proper extent of operation.

The system is needed, which can feedback the result to related team after monitoring selected CRV from online or offline, and make a report. This is called IOW (integrity operating window) system. The concept of IOW system is representing certain limit to operation guidelines of measurable operation variables (Fig. 2). With IOW system, perception of process change can assure equipment reliability by getting feedback and report earlier than condition change is recognized.

Fig. 1 Steps of Development of Corrosion Control Document (CCD)

Fig. 2 Concept of IOW system ("Establishing Integrity Operating Windows (IOW's)", John Reynolds

If it runs out of fixed value, it can make problem like devolution or corrosion of material. We developed CCD as 4 steps. Important content of each step is following.

2.2.1. Step 1: Building a MSD (Material Selection Diagram)

Considering corrosion and economic efficiency, each equipment of oil refinery and petrochemical plant is made from different material. For instance, carbon is used in upper portion of A-column in CDU process, but lower portion is made from Hastelloy that hardly ever corrode. Compared to upper portion, such corrosion materials as asphalt and coke in lower portion are comparatively less pure, and these flow more through lower portion than through upper portion. Therefore, using existing materials, carbon, may cause several problems.

At first, we drew the Process Flow Diagram of each process. Generally, one PFD includes following information.

- · *Process Piping*
- · *Major equipment items*
- · *Control valves and other major valves*
- · *Connection with other systems*
- · *Major bypass and recirculation streams*
- · *Operating data (temperature, pressure, mass flow rate, density, etc.)*
- · *Process stream names*

On completed PFD, We make MSD which reflect all the past histories of material use. Based on design data of corresponding process and the past history, we indicate such the qualities of material as Cr, Duplex, 400SS, Hastelloy, and 300SS on PFD. While

making MSD, we marked the spots where the probes have already been installed and where we need to install other probes.

Following Fig. 3 shows MSD of A-column in CDU process.

2.2.2. Step 1: Building a MSD (material selection diagram)

After building a MSD, there's a need to select a Damage Mechanism. The corrosion materials and the mechanism should be found during this step.

First of all, the design and the operation data should be collected, and then check up the temperature, the pressure and heat treatment conditions from collected data. Once finishing collection data, it's necessary to collect the corrosion material data of each steam. What should be collected is sulfur, chloride ion, hydrogen sulfide, ammonia and so on. The followed step is selecting a DM related to an equipment or each part of equipment refer to API(American Petroleum Institute) 571.

We selected 9 DM out of 19 DM on API regarding to possibility of occurrence at current processes, Creep/Stress rupture, Fuel ash corrosion, Oxidation, High temperature sulfidation, Naphthenic acid corrosion, Hydrochloric acid(HCL) corrosion, Ammonium chloride(salt) corrosion, Wet H2S corrosion and Ammonia stress corrosion cracking.

There are the specific explanation of selected 9 DM, affected materials, main factors, the methods of prevention & alleviation and the ways of inspect & monitoring.
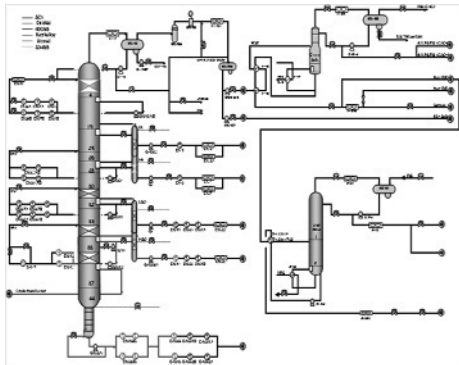
Fig. 3 MSD of CDU process   Fig. 4 DM of Equipment (API 571)
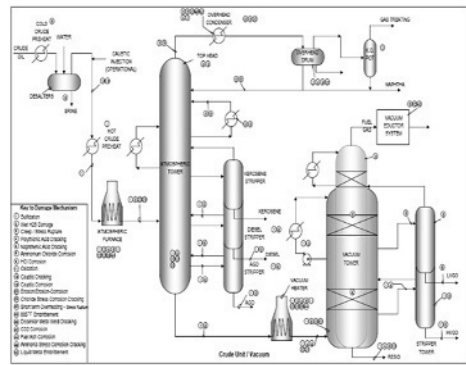
We describe the DM according to equipments and more focuses on parts of equipment. For example, the upper part of A-column has HCL corrosion and Salt corrosion. So are that the middle part has Salt corrosion as well as the lower part has Naphthenic acid corrosion and High temperature sulfidation.

2.2.3. Step 3: CRV (Critical Reliability variables) Selection

After selection the DM, the next step is choosing key parameters that influence DM severity as Critical Reliability Variables.

Main variables are PH, Chloride ion, Hydrogen sulfide, ammonia, TAN and etc refer to Best Practice, API 571 and API 580 & 581. As it's already mentioned, the upper part of A-column has HCL corrosion and Salt corrosion. The influencing factors on HCL corrosion are Water PH, Water CI-, Water Fe, Water H2S, Water NH3 and

Temperature. In addition, measuring temperature has to be always done, whereas the others could be inspected 3times a week.

2.2.4. Step4: Operating guideline proposal

The last step is to introduce a CRV operation guideline which has no influence on the reliability of the equipment (references: API 571 and API 580 & 581, Chemical Vendor Guideline).

Fig.5 illustrates the final developed CCD. It is essential to set the boundary of main variables which affects each equipment so as not to damage reliability.

For instance, PH level should be 6~7 at CRV of HCL corrosion, Cl⁻ must be lower than 30ppm, so is that Fe is lower than 1ppm. This guideline is the basic data of an IOW (Integrity Operating Window) system. The IOW system automatically makes a report every time each variable goes beyond their upper and lower bound. To keep the reliability of the equipment, it's required to pay special attention to the variables which goes beyond their limits too many times.

Fig. 5 Corrosion Control Document

Fig. 6 Relationship among CCD, IOW, RBI

3. Result and Application

3.1. Application to RBI

CCD development becomes a basic data for RBI(Risk Based Inspection) which deals with the priority orders, methods of equipment examination as well as cycle by conducting quantitative analysis about equipment troubles and accidents caused by corrosion.

Unless there's enough data of corrosion factor, it is not possible to make full use of RBI which aims control obsolete equipment. Therefore, there's a need to apply CCD DB to RBI (Fig.6).

That would be able to figure out quantitative risk due to corrosion. Moreover, it could not only improve profit where companies have the methods of risk inspection to reduce the risk through minimum costs and endeavor, but also enhance the safety of process equipment through developing decision-making guideline based on RBI system which has different inspecting cycle and methods refer to the priority order of equipment risk.

3.2. Risk Monitoring and Process Simulation

We could achieve some side effects such as prevention against fatal accident which causes casualties and poverty damage by development of risk monitoring and analysis

algorithm. If we hope to development this system, we should define the important factors which bring about a lowering efficiency such as the continuous flames on pumps, valves, generators, ventilators or heat exchangers.

It's possible to control equipments with significant effects through process simulation. For example, the development of a corrosion scenario and simulating make an accurate estimate of risk linked with probability of accident. Furthermore, it makes reducing unexpected interruption of plant. On top of that, as we've pointed out, the optimal operation guideline with estimation of high risk brings a number of advantages such as reducing unexpected shutdown and downtime, boost output, improvement in quality, executing the optimum equipment and innovation in safety management system.

4. Conclusion

This study focuses on the development of CCD to enhance reliability of the whole process equipment as well as to operate smoothly in chemical corrosion. CCD is developed by building a MSD, selecting DM, CRV and operation guideline proposal. We sincerely believe that CCD could improve the reliability and the safety of the whole refinery industry, and so is sharing information with each department of company. The further step is IOW system could be coupled with RBI system. Besides RBI, we could reduce the risk through simulating process and risk monitoring.

Especially, building CCD is terribly important as the first step for the reliability of equipment and the safety of a plant. And IOW, RBI and Risk Monitoring become the fundamental data of subsequent research.

Finally, we could improve the reliability of the whole refinery, petroleum industry by applying developed CCD to FCC (Fluid Cracking Catalysis), HOU (Heavy Oil Upgrade), LPG and aromatics process due to CCD based on CDU process.

Acknowledgments

References

G.H. Koch, M.P.H. Brongers, N.G. Thompson, Y.P. Virmani, J.H. Payer, "Corrosion Costs and Preventive Strategies in the United States", FSP & SFPE, 2003.

M.S. Ray, M.G. Sneesby, "Chemical Engineering Design Project: A Case Study Approach", 2nd Edition, Gordan and Breach Science Publishers, ISBN 9056991361, 1998.

API(American Petroleum Institute), "API RP 571 Damage Mechanisms Affecting Fixed Equipment in the Refining Industry", 2003.

J. T. Reynolds, "Risk-Based Inspection - Where are we today?", CORROSION 2000, NACE International, Paper No. 00690, 2000.

J. T. Reynolds, "The Application of Risk-Based Inspection Methodology in the Petroleum and Petrochemical Industry", ASME, PVP-Vol. 336, p.125, 1996.

1845

Methodology of Pipe and Equipment Layout for On-Shore Oil & Gas Industry

Luiz G. Persson[1], Flávio B. Santos[1], Cesar A. C. Tavares[1], and Arnaldo E. de Andrade[1]

[1]*Chemtech – Departament of Pipe design, Rua da Quitanda, nº:50, CEP 20011-030, Rio de janeiro-RJ, Brazil.*

Abstract

Traditional process engineering methods in the petroleum industry are based on the experience of engineers' and the well-succeed known cases. The maintenance of the facilities and design methods broaden the scope of traditional methods, but focus primarily on short time and simple construction ways while ignoring the best economic pipe and equipment layout. This paper presents a work method developed during the project of a new refinery located in the northeast of Brazil called RNEST licensed to Chemtech Company. Matching the study of the basic project with the petroleum chemical process it is possible to achieve a new approach which considers inputs from both maintenance and construction design attending the normative security rules. This approach was developed from the study of a small facility of RNEST, the caustic treatment unit, where the reduced number of equipments and pipes made the method possible to be done in a fast an understandable way. The preliminary study involves a review of the general refinery plant and the chemical processes of the facility, that way it is able to rearrange the equipment layout in order to optimize structures and pipe layout. Combining different disciplines and analyzing every sub product inside the unit, the pipe modeling work becomes more efficient and it improves the performance of the system. Together, pipe and equipment analysis can provide insight into the economic field, physical space and sustainability of the industrial process. The proposed method is broadly applicable to assist decision making in petroleum and chemical engineering tasks.

Keywords: refinery, equipment layout, pipe arrange.

1. Introduction

Plant layout design plays an important part in the design and engineering phases of any industrial facility. This paper discusses the role and responsibilities of the plant layout designer, provides advice on how to use project data and mainly develop a methodology of equipment arrangement design and piping layout design.

2. Methodology

In almost any engineering project gathering the basic information is crucial for a successful work. However, there are some unique features in each project, especially how this information is supplied. In a refinery, the pipe layout design requires a good number of personal resources for its full achievement, therefore, a good study at the beginning of the project can minimize redundant work and save precious hours in some cases, as if any equipment has to move from its original position because of bad

equipments location study or save space through a compact layout of all units and their interconnections.

For the conceptual project the first step is analyze the plant production capacity and also parameters there are directly connected with it. Some o those factors as electrical power available in the location, natural resources, water supply and qualified workers, rarely are taken into account, but for a good venture they are extremely important. There is also the local logistical capacity for import and export goods and the frequency of local predominant wind that certainly will influence in the final result.

In the case of pipe arrangement, the process plants (on-sites units) are the most important variables to be analyzed. In those plants are the pipes with high complexity of connections within a short area for construction, this fact requires a detailed analyses of equipments displacement in order to make a good pipe arrangement. In the off-site are the main distribution pipes, they are much bigger than the pipes placed on site, but these pipes have less complexity and their only function are connect the plants of on-shore oil and gas industry. In most of the projects the great challenge of the piping and equipment layout team is to make a layout preventing all piping difficulties such accesses to operation, predict maintenance area and plan access to build the pipes. Commonly those topics are brought into project in the pipe design phase but they could be studied along the equipment design phase in order to perform a better pipe layout and optimize the facility area.

The beginning of the main project is the study of the process plants stats with the descriptive memorials (MD's), supplied by the process team. It is also necessary the analyses of the equipments data sheet, together, these documents describe preliminary equipment sizes and inlet and outlet products and flows. The MD's describe the whole unit process, the function of all equipments and the main fluids pressure and temperature at each process stage. The equipment layout interacts directly with these documents; the MD's shows the fluid flow work and show the order that the fluid must pass through the equipments making easier the layout task. Many systems are described by this document and their interactions with others systems such blow-down, pup-out, compressor, water distributions and others.

As explained, the MD's gives a macro view of operation functionality of the refinery, a deeper analyzes should be made using the Process Diagram. With these documents information adding the reserved area for the plant, the preliminary layout of equipments can be done. A well made work at this level means a compact layout reducing the needed area with the most important pipes previously planed. This means space and pipe material economy.

```
┌─────────────────────────────────────────────────────────────────────────┐
│  ┌──────────────────────────┐       ┌──────────────┐      ┌─────────────┐ │
│  │ Equipments + Preliminary │       │   Space      │      │ Conceptual  │ │
│  │ Process Flow Diagram +   │   +   │ availability │ ───▶ │ Project –   │ │
│  │ Specification + Data Sheet│      │              │      │ Basic Layout│ │
│  └──────────────────────────┘       └──────────────┘      └─────────────┘ │
└─────────────────────────────────────────────────────────────────────────┘
```
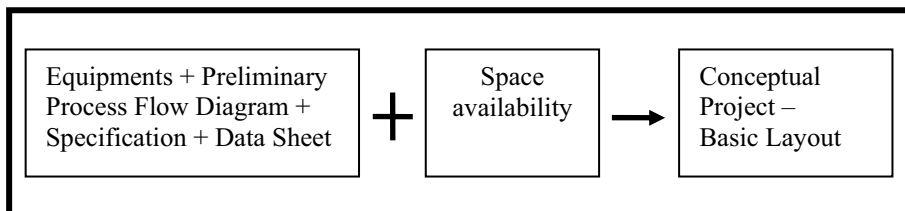
Besides the technical demand, an engineering project should also respect the security and environmental regulation rules. A good equipment and pipe layout should provide fewer costs within maintenance and constructions and also ensure less cost with insurance policies.

Frequently the main company of the venture has its owns regulations and standards for an initial approach and support for the project, some regulations and standards may belong to the company but there is others who was created by a regulation institute, mainly they are set on the contract and must be followed by the project company.

Most of the times the norms are preventatives and their main purpose is orientate the designer, in some cases the designer should choose different types of solution in some appropriate occasions.

2.1. Equipment study

The equipments are displaced at certain logic order to minimize piping materials and the area needed to build the plant. However, its elevation must be considered too. The information needed to study and calculate the equipments elevation are found at their data sheet, pipe and instrumentation diagram and pipe list, which is a document that lists all pipes and their characteristics such as fluid temperature and pressure.

This study has different objective from the pipe rack study. Here, the objective is to indentify the pipes that require certain elevation to work with the equipments and by this work one can determinate the elevation necessary to all pipes to work well. Since it's impossible to think at the same time in all the structures, valves, buildings and equipments that maters in the elevation of a single equipment a lateral view pipe sketch of the equipments and the most important pipes that connects it is a good way to study its elevation.

This kind of pipe sketch works in a very similar way to the pipe sketch presented previously. It's important that this sketch incorporates information about equipment nozzle elevations, building near the equipments and equipments that works together with the studied equipment, with that ready then the important pipes must be drown. The designer should also consider that some equipment has a minimal elevation defined by the process team, and since all equipments are connected they need to respect a certain elevation between then. For example, one pipe that drains fluid at low pressure from one vase or tower and head to the pump must have the drain point higher than the pumping point; in other words, the fluid must flow by gravity.

As said before the basic project or the process team determines a minimum elevation for the equipment, and this study is responsible for determining if this minimum is enough or not. All structures and buildings in a plant are always associated with a certain equipment, defining its elevation is important, so the project of these structures can be started by other teams in the project.

2.2. Pipe Sketch

The complexity and the high number of pipes that are required into a petrochemical plant make too complex analyzing one system or determinate the structures and buildings dimensions. These analyzes must include all or almost all the pipes that pass at the studied area and depending on the quantity and the complexity of these pipes the designer imagination may not be sufficient to consider all variables. To make these analyzes we suggests a methodology presented in this paper as the pipe sketch, which consist to draw the pipes, on the layout arrangement drawing, in 2-D representing them by a line. This drawing can be simple and not precise with their length.

2.3. Pipe-rack Dimensions

In a refinery the number of pipes is too high and the pipe rack is constructed in order to support almost all pipes and it's also a way to organize the pipes grouping them all. To determinate the necessary pipe-rack dimensions and avoids area and material loss the pipes passing trough the pipe-rack must be studied. This study must consider the critical section of the pipe-rack, which means the section that has more pipes to be supported; to determinate this section the designer must make a pipe sketch of all pipes that he intends to pass trough the pipe-rack. After that the designer must locate all pipes in the section and calculate the space necessary .On our case study we added 30% of free space for further plant improvements or unanticipated pipes. When finished the designer must decide which way he can build the pipe rack, high and thin or thick and low, this choice depends on the available area.

This is for sure an arduous work, but avoiding this study make the chances of failure grow. An overestimated pipe rack means material and area loss and an underestimated one means a probable project failure.

2.4. Maintenance and building analyzes

With the basic layout done, some important points must have special attention at this moment to improve the quality of the project. To complement the equipment layout the point showed below have to be studied:

- Access to valve and instrumentation operation
- Free area to equipment maintenance
- Platforms to access and operate all needed equipments and instruments
- Emergency exits
- Area and accesses for building hoist and other building equipments

These topics has to be thought together with the all others project topics, but its essential to have an specific time just to check if these topics are correctly applied in the project at each system, each equipment and each valves control station.

The first system to be analyzed should be the pressure relief system; most of PSV's (pressure safe valves) have to be above the flair line (the header of all relief pipes). Since the flare may be at the last level of the pipe-rack it's convenient to place a platform along all pipe rack top to operate and place the PSV's control stations. These platforms can be connected with other buildings and equipment platforms, in order to make easier the accesses to the higher level of many buildings. Depending on the available area, constructing platforms at the side of the pipe rack at medium level can be a good solution since the designer can place control stations, instruments and valves that don't have space to be placed at the floor on these platforms.

Security is very important to any industry, in oil and gas industry isn't any different; respect the security rules are many times very expensive, but not attending these rules can cost much more, so at any project one must attend all security rules that apply in each case.

Access in all plant frontiers can be useful for building, maintenance and secure the plant. These accesses are important to be large enough to support the transportation of the big equipments such as reactors, towers, compressors, pumps and vases.

In the pipe design of the plant it is important to think if the pipe project is possible to be build verifying the equipments already located and buildings near the pipe. Some pipe project has welds that are impossible to be done, so the project must have these cases

reduced to the minimal possible avoiding remaking the pipe project or difficulties in the building field.

3. Case Study

As it could been seen on the Abstract, the methodology above was developed through a refinery project in the northeast of Brazil. Among all problems solved in the Project, the main focus was use a reduce area and make a compact pipe layout. There are many ways to make a compact layout but many of them may not be fully functional, resulting in difficulties for building and operate the plant.

The refinery project is ambitious and its main objective is to take advantage of the market available at the Brazilian northeast region and supply the Brazilian demand (that currently import diesel at higher prices, mostly because of the transportation and refining cost which increase the final product cost), maintain a strategic position at European market, which have a high diesel prices and a big demand, and finally to attend international quality requests such as low sulfur and other pollutants.

In this case study, we will present two plants of the refinery, were we have change the basic layout in order to have a performance improvement. The main alterations from the basic project done were: pipe rack dimensions, new access to the plant and equipment location.

The case study starts with the MD's and process diagram study and with the basic project on hand. Then the layout was modified to a more compact form respecting the norms and regulations required by the owner. The frame below (Figure 1, Figure 2, Figure 3, Figure 4) shows the after and before layout for the plants studied.

Together with the layout comes the pipe-rack dimension study from both plants. In unit I within the pipe sketch we determinate that the pipe-rack was under provisioned for the amount of pipes it should support. The total length was 24 meters divided in 4 levels. Taking the most critical session of the pipe rack from the pipe sketch we could set the dimensions to 50 meters at least without consider possible expansion or unanticipated alterations that could require more space.
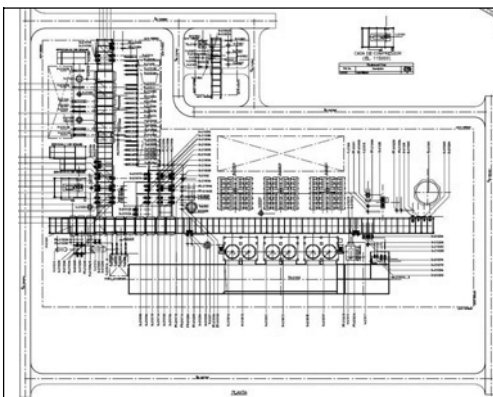
| Figure 1 –Basic equipment layout Plant – Unit I | Figure 2 – Final equipment layout Plant – Unit I |
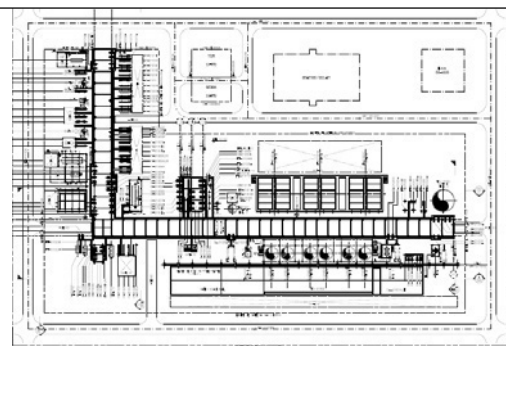
Figure 3 – Basic equipment layout Plant-Unit II	Figure 4 – Final equipment layout Plant-Unit II
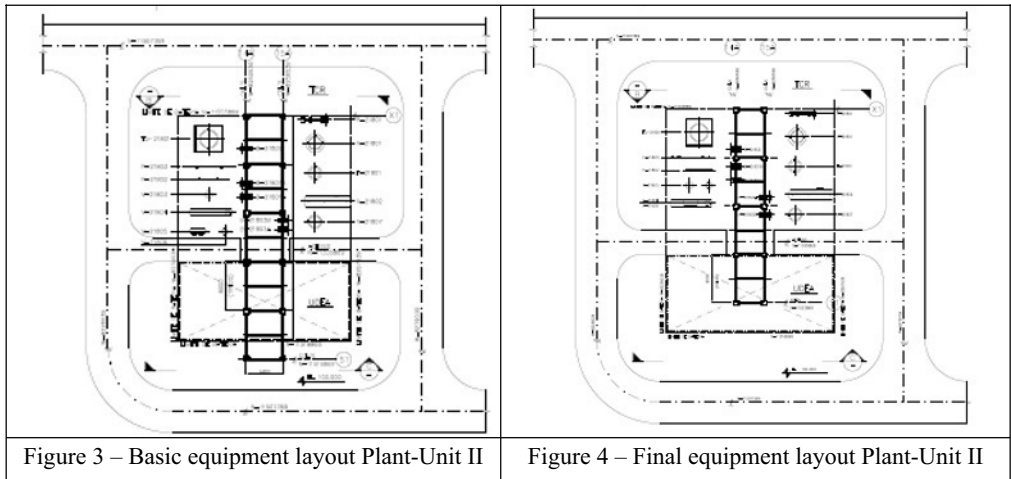
The operation platforms for pressure safety valves are located in the top o the pipe-rack. This way the pipes that contain valves don't need to reach the floor and find space. With this platform pre-projected one can design a better pipe arrangement wasting less pipe components. A similar study was done for each equipment in order to develop a compact pipe layout with easy access for maintenance and operation.

4. Conclusion

It could be seen that plant layout design plays an important part in the design and engineering phases of any industrial facility. The development of equipment arrangements and piping layouts for process off-shore industries offers an opportunity to demonstrate technical ability along with a creative talent and common-sense approach to problem solving. Process facilities must be designed and engineered within extremely short schedules while adhering to maintenance, safety, and quality standards; moreover, the design must take constructability, economics, and operations into account. Although the tools to achieve these goals are changing from pencil and paper to computer graphics terminals, the responsibilities of the plant layout design remain the same.

Each plant layout develops an individual layout philosophy. Although conditions (client specifications, schedule constraints, and availability of information) may change significantly among projects, if the design style philosophy remains the same, the probability of a successful project is very high.

The main objective it is to keep in mind all possible variables to optimize the layout design project.

References

Z. R. Hunt, 1993, the Basics of Plant Layout Design, Process Plant Layout and Piping Design, PTR,

Y. P.C.S.Telles, Instalationd and Construction Industrial Dsiposal, Industrial Pipe Design, LTC, 10 Edition

X. ASME, 2006, Standards for Piping Components , Process piping, Code for Pressure Piping B 31.3, HIS.

1851

Chemtech's Energy Efficiency Methodology

Renata Machado[a], Andreas Hahn[a],Rafael Teixeira[a], Flávio Waltz[a], Valter
Souza[a]

[a]*Chemtech – A Siemens Company, Rua da Quitanda, n.50, Rio de Janeiro-RJ, CEP
20011-030, Brazil*

Abstract

There would be no lack of energy if all the existent technologies were applied for a
more efficient energy use, said Greenpeace – Brazil. Aiming at an energetic
sustainability of his clients, Chemtech developed a methodology that primes for the
conscientious use of energy, contributing to the reduction of industries' energy
consumption. This methodology is based on the traditional energy audit, consisting of
an initial assessment, data survey and measurements, resulting in actions to reduce
overall energy consumption, allied to a differentiated implementation. The audit
embodies a tariff analysis, correction of power factor, pre feasibility study of
mechanical and electrical measures and process analysis. A conceptual project is
elaborated on selected measures (i.e. automation, production, electrical installations and
utilities) after a detailed feasibility study. Differing from a traditional audit, a client's
energy management team is established aiming the development of an internal control
of energy, with the implementation of an energy efficiency awareness program for all
employees and controlling the energy consumption through established energy
indicators. Another differential of Chemtech's work is the approach used to develop the
audit. Each milestone of the audit consists of several steps, each step used to augment
the consistency of gathered data and refine obtained results through recurring meetings
with the client covering specific topics. The next step is only initiated after the client's
acceptance of data and results pertaining to the previous one. This approach raises the
consistence and acceptance of the developed work, like a helix, guaranteeing the overall
development efficiency, so as the consistency of the project.

Keywords: Energy efficiency, energy audit, energy management.

1. Introduction

Brazil have been making efforts on energy conservation since 1985 with the creation of
the Programa Nacional de Conservação de Energia – Procel (National Energy
Conservation Program), focused on electricity. This search for energy efficiency
culminated with the creation of the Programa Nacional da Racionalização do Uso dos
Derivados do Petróleo e do Gás Natural - Conpet (National Program for the
Rationalization of the Use of Oil and Natural Gas Derivatives) in the year of 1991.
Nowadays, with increasing concerns about the future of energy availability, energy
supply security has been spotlighted in the political debate around Brazil. Not counting
the global climate change, caused by the increase of green house gases on earths
atmosphere, in detriment of ecosystems, causing stronger and more frequent storms,
melting of glaciers, heat waves and droughts. Questions like economical and
populational growth, beside energy intensity are some of the key variables responsible
for the increase of energy demand.

In this context the energy efficiency activities become usual in industrial plants, buildings and even homes. Nowadays most of brazilian energy conservation companies (ESCOs) attained plenty knowledge about how to develop energy audits and implement selected measures on their clients' plants. These are the main activities developed on the brazilian energy efficiency market so far.

In spite of the high initial impact of pure technical energy efficiency measures, these lose part of their effectiveness along the time if no constant monitoring of their performance is done. For this reason and that the human factor influences profoundly the energy performance of a company, Chemtech opens a second action front through the energy management, obtaining synergy with the proposed technical measures.

2. Objective

This work intends to present the energy efficiency methodology develop by Chemtech, focused on two major fronts: one aimed on the technical and process approach. The other one aimed on the development of a continuous client side energy management. Both fronts are associated with an integration methodology widely used by Chemtech, directed to the client's feedback.

3. Energy Audit

3.1. Workshop

The first step for the development of an energy audit consists of a workshop, involving client's employees of different interest areas (automation, maintenance, operation, among others). The objective of this first workshop is to gather initial information for a preliminary energy report. The survey is mainly focused on:

- assessment of the production vs. energy flow on the site;
- identification of the main energy consumers on the site;
- identification of the main energy loses;
- identification of bottlenecks;
- determination of the energy audit's foci.

Most of the time the client's employee's opinions, due to their large experience on field activities, provide information regarding energy consumption reduction, being of great value during the planning of the next steps.

During this first contact a communication plan between the companies is established, providing an efficient link able to supply the communication needs during the project development.

At the end of this stage the team should be able to tell the main bottlenecks of the production processes and some savings opportunities.

3.2. Preliminary Energy Report

During the development of the preliminary energy report three activities are developed: on field assessment, assessment of maintenance needs and assessment of operational needs.

During on field assessment the main energy consumers are verified. Internal and external lighting systems, compressors, pumps, ventilators, motors and other equipments with energy consumption reduction potential are analyzed. Last but not least, the utility consumptions are focused: compressed air, HVAC and water. An overall analysis of the electrical structure of the site is also initiated.

Most of the time, the lack of maintenance of some equipment, due to the lack of maintenance control, turn these obsolete, inefficient or inadequate for use, causing a

higher energy consumption than necessary. In this case the use of good maintenance management practices leads to the reduction of energy and utilities consumption, sometimes with unexpected high savings.

The assessment of operational data is an essential activity for the further development of the energy audit. At this stage the energy costs, production and consumption reports and process documents are evaluated.

3.3. Energy Audit

During the development of the energy audit the on field assessment is terminated and an evaluation of the main potential energy savings is made. For each potential measure the monetary savings are calculated and a technical and economical feasibility study is developed. Then a second workshop is necessary to define, together with the client, which measures should be detailed for implementation based on the results of the feasibility study. At this workshop the proposed measures are classified in short term and long term projects, last ones leading to conceptual projects.

At the end of this stage the client has a definition with all selected actions to be implemented associated to their energy savings potential.

3.4. Short Term and Conceptual Measures

The short term measures are immediately implemented by the client with support of Chemtech's energy efficiency team, to assure the expected savings.

For the long term measures conceptual projects are elaborated aiming the development of future basic and detailed projects for implementation.

4. Energy Management

In spite of the great initial impact of pure technical measures, at long term, this approach loses his effectiveness if no further follow-up is undertaken. Since the human factor is essential for the continuity of the implemented measures, Chemtech developed an energy efficiency management methodology to guarantee the endurance of implemented technical measures and the development of new ones, and also to develop a sound client-side energy management structure. This methodology involves the assessment of the clients' administrative and communications structure, involvement of different sectors of the companies, application of specific questionnaires, training of the members from the newly created or already established energy management staff to improve their energy efficiency and management skills, including monitoring, targeting & reporting, and the creation of an energy efficiency awareness program for all employee levels. Key performance energy indexes to manage results are customized attending the client's needs, based on traditional indicators and on the cumulative sum of differences methodology, usually used in the statistical process control, but now applied for monitoring the overall energy consumption performance between well defined limits.

4.1. Methodology

Chemtech's energy management methodology is a blend of different methodologies published mainly by Natural Resources Canada and ETSU (UK) and years of experience of its energy efficiency professionals. The core of the methodology relies on an awareness program implemented together with monitoring & targeting of key indices defined together with the client.

Experience has shown that several steps are necessary for the implementation of a successful energy management program, each of them discussed below.

4.1.1. Joining the Players

This step joins the players necessary to maintain a sound energy management program. First of all a formal commitment from the senior management is necessary. No formal commitment usually leads solely into an implementation attempt, with no expressive returns. Also an energy management sponsor has to be defined together with an energy management team. An energy management leader is assigned with powers and authority to develop company's energy policies.

4.1.2. Identification of Opportunities

Together with technical opportunities, pointed out through an energy audit, human resources have to be evaluated, mainly through interviews and questionnaires. The results obtained from the human resources' feedback will indicate needs for awareness or specific training. Human resources are essential for the success of any energy management program; special care has to be taken during this step assuring that most of the employees needs for improvements are listed.

4.1.3. Establishing Targets

Targets are set not only for energy consumption levels, but also for the awareness and necessary skills of the employees. Human resource needs have been unveiled during the identification of opportunities and at this step a new level of knowledge and awareness is defined to be reached in a certain amount of time.

Energy consumption indices are determined together with the client and these will be monitored to be maintained between well defined limits. One of the main tools used during this step is the cumulative sum of differences methodology, derived from the statistical process control. Energy consumption and production data are correlated and evaluated; the best historical performance is selected and set as desired target. Since this energy consumption performance already had been reached once, no objections can be made about it. Later better energy performances will be achieved and set as new targets during the program evaluation.

4.1.4. Communications Plan

Mostly not given value to by technical personnel, communication is an essential part of energy management, since most results proceeding from it are achieved from the company's employees.

At this step existing communication lines are assessed, target audiences are defined, messages developed and communication tools identified or created. Sometimes energy efficiency events are planned at this point. A great amount of time is invested in this step, since a communication failure usually leads to poor energy management results.

Client's communication department is to be closely involved, working together with the energy management team, designing strategies for the following months.

4.1.5. Training Plan

In parallel with a communication plan a training plan is developed based on the results of the identifying of human resources opportunities and needs. The plan is designed to achieve the targets set for employee's skill and awareness levels and the training can be delivered in house or by third party. Usually the energy management leader, together with the energy team, defines the training needs, but not necessarily delivers the courses themselves. The program evaluation indicates if the desired skill and awareness levels have been achieved.

4.1.6. Energy Management Implementation

The energy management program is formally implemented only after the steps above had been fully developed. The implementation is preceded by a communication

campaign raising employee's awareness and preparing the company for the monitoring and targeting to be initiated. At this point usually a formal commitment from senior management is published.

4.1.7. Program Evaluation

After implementation the energy management program has to be constantly evaluated, adjustments have to be made to targets and to the overall management plan. New interviews and questionnaires help to evaluate the success of delivered training courses. Increasing awareness leads to more feedbacks, which, together with the necessary adjustments, creates a continuous improvement loop.

Energy performance reports are send at defined intervals to decision makers and to energy customers allowing them to follow the consumption trend and make necessary adjustments to achieve the desired consumption levels. Deviations from desired consumption levels are investigated and adjustments made to achieve targets or new target levels are set.

4.1.8. Sustainability

To guarantee the sustainability of the energy management program some activities are necessary to maintain a high level of response.

New employees have to attend training sessions about energy efficiency and the company's energy management program basics to level up with the rest of the client's employees. Recurring communication messages have to be developed to reinforce the awareness, taking care not to saturate the audience. Success has to be shared recognizing success and pointing out exceptional achievements.

Last, but not least, the whole program has to be evaluated from time to time, being adapted to be up-to-date with the company's needs.

5. Chemtech's Integration Methodology

During the whole development time of preliminary energy reports, energy audits and even the implementation of an energy management, aiming the best roll out of the project management, Chemtech applies its integration methodology.

This methodology consists of recurring meetings with the client during the development phases discussing obtained data and results obtained so far. At each meeting existing data and results are reevaluated and necessary corrections are made, like a spiral, each beginning "coil" of the spiral representing a meeting and the "height" of the spiral representing the growing knowledge and accuracy.

Chemtech's integration methodology keeps the client informed about the development at different phase levels, obtaining his necessary confidence, participation and support, managing client's expectations and avoiding unnecessary rework at the end of the project. This approach has been largely appreciated and praised by Chemtech's clients.

6. Conclusion

The traditional energy efficiency approach through a purely technical energy audit with further implementation of energy saving measures, although with high initial impact, loses effectiveness over time if no control is undertaken to assure continuity of the implemented actions.

Although mostly neglected by technical personnel, communication is a fundamental part of the success of any energy audit or energy management program. Chemtech's integration methodology is designed to improve the communication with its clients during development and a sound communications plan assures best results through an implemented energy management program.

Allied to an energy management program the results obtained through technical measures can be maintained and even increased due the interaction with company's human resources. The employee's skills and awareness can be increased leading to a sustainable energy management.

References

Saving money through energy efficiency – A guide to implementing an energy efficiency awareness Program, Natural Resources Canada, 2004

Guide 112 – Monitoring and Targeting in large companies, ETSU, 1998.

New Model to Determine Fracture Gradient in Ultradeep Water

Clovis D. Ferreira,[a] Wilson da Mata,[a,b]

aPós-Graduação em Ciência e Engenharia de Petróleo, Universidade Federal do Rio Grande do Norte, Campus Universitário, Natal-RN 59078970, Brasil
bDepartamento de Engenharia Elétrica, Universidade Federal do Rio Grande do Norte, Campus Universitário, Natal-RN 59078970, Brasil

Abstract

The development of the offshore drilling in ultra deep water requires additional care in function of some critical points. In the beginning of drilling, just below the bottom sea level and the transition zones where pore-pressures tend to the formation fracture pressures. With the increase of the depth of water in relation to depth of sediments below the bottom of the sea, it happens the reduction of the operational window that it is the available space to work with the density of the drilling fluid, hindering the progress of the drilling. To minimize the problem, it is proposed in this work a model that uses the superior limit of the operational window (fracture gradient) equal to the value of the leakoff test available of the area in the depth considered to use as maximum value for the gradient of mud. The calibration adjusts the fracture gradient made calculations to the leakoff test. In this situation, it is been in favor of the safety, being avoided fractures of the formation with loss of circulation or kick. The proposed model is a geometric correlation that uses values of pore pressure gradient and overburden gradient of another indirect model and executes the calculations using a relationship in which the involved parameters vary in the same proportion.
The application of the proposed model implies in costs reduction in the definition of the casing shoes depths, in the casing sizing reduction, in a better cementing design and also in a more adequate mud weight for drilling all the well phases. The economic return is more significant in scenario of ultra deep water.

Keywords: fracture gradient, overburden, pore pressure

1. Introduction

The progress in the search for news offshore ultra-deep waters oil reservoirs requests from petroleum industry high investments in new methodologies and new technologies. For that, there is a growing need in the development of special tools to drill safety oil wells.

One of the challenges to drill these specific oil wells is the project of a drilling fluid which must have a hydrostatic pressure to balance the low fracture gradient normally founded in these oil reservoirs. The critical points are the beginning of drilling, just below the bottom sea level, and the transition zones where pore-pressures tend to the formation fracture pressures. In both cases, there will have a narrow margin pore-fracture of working densities for drilling fluids. Bigger is the sea water depth, less is the margin of pore-fracture pressure, limiting continuous drilling operations until reach the oil zones. By consequence, normally more complex well designs are needed (many casings and many casing shoes).

The well knowledge of pore-pressures and fracture-pressures defines operations risks to avoid column prisons by differential pressure, formation damage due to drilling fluid invasions or fluid filtrates, formation fracture and bad stability of well walls.

Several models exist for prediction of fracture gradient and normally they can be used in offshore oil well design. The fracture gradient can be obtained through direct or indirect methods.

The direct methods are accomplished through measurements directly made in front of the interest intervals (well tests). For this reason, they are usually expensive because they use rig times, in spite of supplying real and trust values. Normally they serve as a data base for fracture gradient pressure calibration curves generated by indirect methods.

The indirect methods use techniques of calculations based on theoretical models or empiric models supplying an estimation of the fracture gradient along all the well depth. To avoid all these problems during an offshore oil well drilling, all the well phases are planned, taking in account estimating values of fracturing gradient pressures for all the well depth, normally using indirect methods based on approximation of leakoff tests (direct reference tests). For this reason, the leakoff test is assumed to be the superior limit of fracture gradient pressure, permitting more economical and safer wells.

2. Proposed Model

2.1. Observation

The proposed mathematical model is an indirect method based on geometric observations obtained from geopressure logs, which gives the pressure as a function of the well depth.

The data base used for this assumption was the one found in the literature (books, articles, field data, etc).

From geopressure logs, which represent graphics of indirect methods, can be obtained proportional relationships between the pore-pressure and the fracture-pressure, both in the same Cartesian plan profile.

The proportionality relationship is obtained when it is considered the differential pressure, for the same depth, to be represented by co-linear, consecutive and adjacent segments where the endpoints of each segment are pressure gradients.

2.2. Base of the proposal

Through an applied geometric relationship in Barison (2008) based on the principle of proportional segments, the equation (1) can be obtained where F/D (fracture gradient) represents the endpoint of segment. In equation (1), F/D is the fracture gradient, S/D is the overburden gradient and P/D is the pore-pressure gradient

$$\frac{F}{D} = 0{,}38197 \frac{S}{D} + 0{,}61803 \frac{P}{D}$$

(1)

Rearranging the equation (1), the equation (2) can be obtained inserting a Calibration factor. This equation represents a linear relationship in the geopressure profile from which can be estimated a fracture gradient approximately equal to the value of leakoff test.

$$\frac{F}{D} = Calibration \cdot \left(0{,}38197 \frac{S}{D} + 0{,}61803 \frac{P}{D} \right)$$

(2)

2.3. Calibration

The calibration is obtained in two steps according to the following conditions:
1st step - This step corresponds calibration to the initial part of the well (where do not exist information) from the bottom sea level to the depth of leakoff test.
For this, it is used the equation (3) in which the value F/D is obtained from equation 1 and X one of the following parameters, depending on the sea water depth:
- Wells located in sea water depth between 300 meters and 800 meters, X =1,000.
- Wells located in sea water depth superior to 800 meters, X =10,000.

$$Calibration = \frac{Leakoff\ Test}{Leakoff\ Test - \left(\frac{Depth}{X} \cdot \frac{Leakoff\ Test}{\left(F/D \right)_{proportional\ segments}} \right)} \qquad (3)$$

where:
Depth - measure since the rotary kelly bushing until the calibration depth before the leakoff test.
2nd step - Calibration from the depth of the leakoff test to the final dept of the well
For this, it is used the equation (4) in which the value F/D is obtained from proportional segments in the same depth of the test. The calibration of this step is a constant value that adjusts the F/D curve, not depending on the sea water depth.

$$Calibration = \frac{Leakoff\ Test}{\left(F/D \right)_{proportional\ segments}} \qquad (4)$$

The proposed model is a geometric correlation that uses the values of pore pressure gradient and overburden gradient from another indirect model and executes the calculation on a simplified form using for this a direct adjust with values of available leakoff test of the area. This procedure is repeated for each depth throughout all the well intervals.

2.4. Application and results

Figure 1 observed in Otatti et al (2008) presents a profile of an offshore directional well 1-GM-1-USA drilled in a seawater depth of 1,493 meters with 5,200 meters final depth. The available data for this well is three leakoff tests and two tolerances to kick tests, which are used to be compared with the literature models and the one proposed in this work.
It can be also verified that the calibration was made using only the depth of 3,100 meters and the curve generated for the proposed model was valid to the others two values of the leakoff test for the depths of 4,350 meters and 4,800 meters.
This evidenced that the values of the pressure gradients obey a proportionality correlation showing that the fracture gradient is a linear function of the proportional segments.

It is observed in Figure 1 that the curve of fracture gradient calculated throughout all the well for the methodology proposal, presented approach values with the prompt values of leakoff test that the gotten ones with the existing methodology.

Figure 1 – Different between profiles of fracture gradient proposed and method existent observed in Otatti et al (2008).

The available data for this well is four leakoff tests shown in the Table 1, which are used to be compared with the model proposed in this work.

Table 1 - Presents the differential pressure between the values of calculated fracture gradient and the proposed model, based on leakoff tests of Figure 1.

Depth (m)	Leakoff Test (lb/gal)	Existent Model		Proposed Model	
		F/D (lb/gal)	Error %	F/D (lb/gal)	Error %
3,100	11.5	11.1	3.48	11.5	0.00
4,350	13.3	13.5	1.50	13.4	0.75
4,800	14.5	14.3	1.38	14.5	0.00

Other example from the literature of Rocha et al (2004) in wellbore stability analysis for a deepwater well drilled in Marlim Sul Oilfield, offshore Rio de Janeiro of Brazil.
The Figure 2 is shown the profile of geo-pressures of the referred well.
It is a profile of an offshore directional well drilled in a seawater depth of 1,205 meters with displacement of 3,210 meters, total vertical depth of 2,903 meters and total measured depth of 5,211meters.

It is observed in Figure 2 that the values calculated for the existing methodology for fracture gradients throughout the entire well, had been very above of the value of the leakoff test that is the referential considered superior limit.

As the sizing of the mud is based on the values of fracture gradient, it means that, problems with losses of circulation had probably occurred caused by the weight of the mud raised in great interval of the well.

With exception only of the depth of 2200 m that the value of the calculated fracture gradient almost equaled with the value of the leak off test.

It is verified that the curve of the fracture gradient obtained in Rocha et al (2004) it crossed the value of the leakoff test in the depth of 1700 m in 0,9 lb/gal and also in the depth of 2870 m where the difference crosses 1,4 lb/gal above the value of the leakoff test, meaning that that value is not compatible to establish the maximum density for the drilling fluid, because certainly it can result in loss of circulation for fracture of the formation.
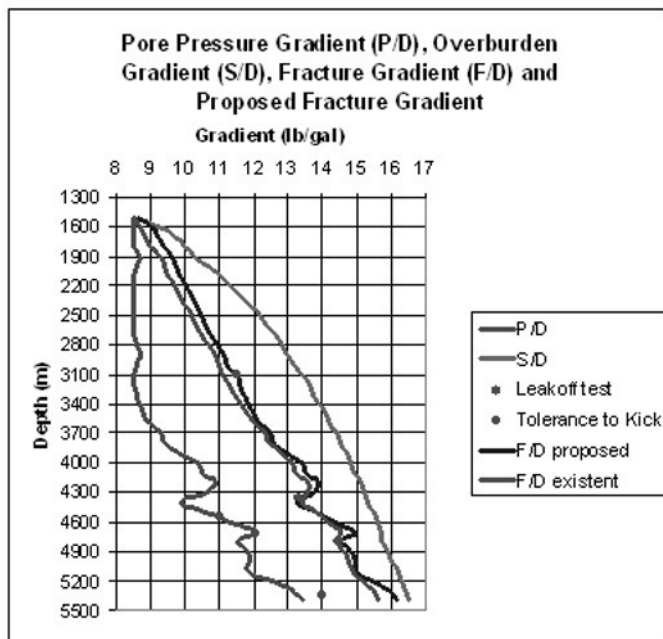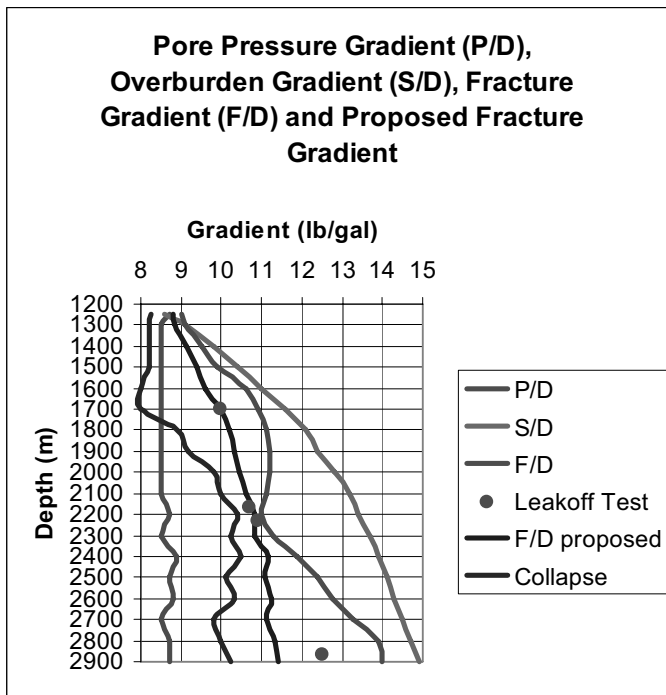

Figure 2 - Different between profiles of fracture gradient proposed and method existent observed in Rocha et al (2004).

The available data for this well is four leakoff tests shown in the Table 2, which are used to be compared with the model proposed in this work.

Table 2 - Presents the differential pressure between the values of calculated fracture gradient and the proposed model based on leakoff tests of Figure 2.

Depth (m)	Leakoff Test (lb/gal)	Existent Model		Proposed Model	
		F/D (lb/gal)	Error %	F/D (lb/gal)	Error %
1,700	10	10.9	9.00	10	0,00
2,170	10.7	11.1	3.74	10.6	0.93
2,230	10.9	11.2	2.75	10.8	0.91
2,870	12.5	13.9	11.2	11.3	9.6

2.5. Conclusions

The model proposed in this work was developed for offshore wells scenarios, since it exists, for the specific area, pore pressure gradient, overburden gradient and leakoff test. It is a geometric correlation that can be used as an alternative form substituting an indirect method to estimate the fracture gradient not depending on the Poisson Ratio of the interval perforated sediments.

The proposed model in this work showed, for the available literature data, that the fracture gradient (F/D) is a linear function of the proportional geometric segments for each depth. For this, a linear geometric relationship allows to establish an estimative value for the fracture gradient with better approach to leakoff tests than those compared with the literature models.

The simplicity of the model proposed in this work, besides can give easier and better results to estimate fracture gradient (F/D), can produce greater efficiency in the calibration with the leakoff test making possible a better exploitation of the available narrow range of pore-fracture. In consequence, a significant cost reduction can be obtained with better well designs, especially offshore ones where the costs are normally very important.

References

L. Rocha, R Andrade. H. Freire, 2004, Important aspects related to the influence of water depth on ERW. IADC/SPE 87218, 15 pages.

H. Otatti, R Lepeleire, V. Ferreira, 2008, Trabalho Final de Curso de Engenharia de Poços da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) On Line in: 26.08.2008. Available in: http://www.gtep.civ.puc-rio.br/petro/movie/pocos/swf/projeto.swf.

B. Barison, 2008, Geométrica: Desenho, Geometria e Arquitetura On Line in 08.08.2008. Available in: http://www.mat.uel.br/geometrica/php/pdf/dg_prop_%C3%A1urea.

1863

Optimization of Steam and Solvent Injection as an Improved Oil Recovery Method for Heavy Oil Reservoirs

Edney R. V. P. Galvão,[a] Marcos A. F. Rodrigues,[a] Jennys L. M. Barillas,[a]
Tarcilio V. Dutra Jr.,[a,b] Wilson da Mata[a,c]

[a]*Programa de Pós-Graduação em Ciência e Engenharia de Petróleo, Universidade Federal do Rio Grande do Norte - Campus Universitário, Natal-RN 59078-970, Brasil*
[b]*Departamento de Engenharia Química, Universidade Federal do Rio Grande do Norte - Campus Universitário, Natal-RN 59078-970, Brasil*
[c]*Departamento de Engenharia Elétrica, Universidade Federal do Rio Grande do Norte - Campus Universitário, Natal-RN 59078-970, Brasil*

Abstract

Currently a resource more and more used by the petroleum industry to increase the efficiency of steam flood mechanism is the addition of solvents. The process can be understood as a combination of a thermal method (steam injection) with a miscible method (solvent injection), promoting, thus, reduction of oil viscosity and interfacial tensions between injected fluid and oil. Solvents are hydrocarbons well known for reducing these tensions and facilitating the production of heavy oil. The use of solvent alone tends to be limited because of its high cost. When co-injected with steam, the vaporized solvent condenses in the cooler regions of the reservoir and mixes with the oil, creating a zone of low viscosity between steam and heavy oil. Mobility of the displaced fluid is then improved, resulting in an increase of oil recovery. To better understand this improved oil recovery method, a numerical study of the process was done contemplating the effects of some operational parameters (distance between wells, steam injection rate, solvent type and injected solvent volume) on cumulative oil production and oil rates. A semi synthetic model was used. Some reservoir data were obtained similar to those found in Brazilian Potiguar Basin and others ones were obtained from literature. Simulations were performed in STARS (CMG, 2007.11). It was found that injected solvent volumes increased oil recovery and oil rates. Further, the majority of the injected solvent was produced and can be recycled. High initial productions achieved by models that use solvent have normally a significant impact on the operation economics, because earlier productions suggest that fluids injection (steam and solvent) can be interrupted earlier. On environmental point of view, it would have a reduction of energy and water consumptions for steam generation, having diminished Green House Gases (GHG) emissions. Also it is important to emphasize that the high oil rates presented by these models can generate an earlier financial return, and this would be decisive for the economic viability of the project.

Keywords: solvent, steam flood, heavy oil, reservoirs modeling.

1. Introduction

Steam flood is an improved oil recovery method applied generally in viscous oil reservoirs. This method consists of injecting heat to reduce viscosity, increasing oil

mobility and facilitating its production. To increase the efficiency of this mechanism, a resource more and more used by the petroleum industry is the addition of solvents.

Diverse hybrid steam-solvent processes are being developed, mainly by Alberta Research Council (ARC), in Canada (Nasr & Ayodele, 2006). Experimental results, numerical studies and field tests suggest that great benefits can be achieved with the addition of solvents to the injected steam, as an increase of oil rates and oil-steam ratios, reduction of energy and also a reduction in water consumptions for steam generation, having diminished Green House Gases (GHG) emissions. Thus, it is indispensable to study the viability of this new technology in heavy oil fields from Brazil, with right adjusts to particularities of each reservoir.

2. Process Modeling

In this work, steam and solvent injection was analyzed through a vertical wells system. A semi synthetic model was used. Some reservoir data were obtained similar to those found in Brazilian Potiguar Basin and others ones were obtained from literature. Simulations were performed in STARS (Steam, Thermal and Advanced Reservoir Simulator) 2007.11 from CMG (Computer Modelling Group).

2.1. Reservoir Modeling

The physical model corresponds to an oil reservoir of 100 m x 100 m x 29 m, on a Cartesian coordinates system (x, y and z directions), as shown in **Fig. 1**. To reduce simulation time and considering system symmetry, the chosen flooding pattern was one quarter of an inverted five-spot, represented by an injector and a producer wells. Some of used data, like rock-fluid properties and heat loss parameters, are represented in **Table 1**. In the bottom hole direction, injector well was completed until the 8th layer, while producer well was completed until the 11th one.
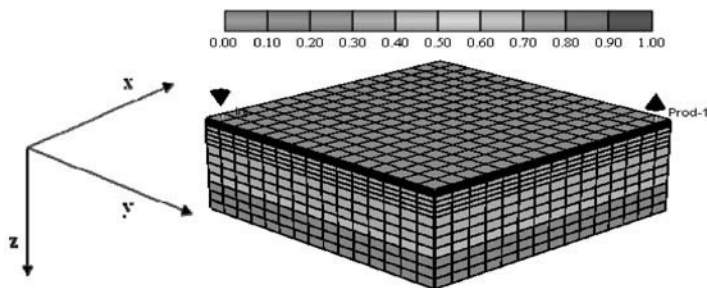
Fig. 1. Reservoir model. Initial oil saturation.

Table 1. Reservoir simulation input parameters.

Item	Value
Reservoir thickness (m)	29
Oil zone thickness (m)	20
Water zone thickness (m)	9
Initial reservoir temperature (ºC)	37.8
Reservoir depth (m)	200
Porosity (%)	30
Vertical/horizontal permeabilities ratio (%)	10

Table 1. Reservoir simulation input parameters *(continuation)*.

Item	Value
Oil viscosity*	1000 cP@37.8°C
Initial oil saturation (%)	70
Irreducible water saturation (%)	30
Steam temperature (ºC)	288
Steam quality (%)	55
Maximum pressure in the injector well (kPa)	7,200
Minimum pressure in the producer well (kPa)	197
Thermal conductivity of rock and surrounding formation overburden and underburden (J/m-s-K)	1.7
Thermal conductivity of water (J/m-s-K)	0.6
Thermal conductivity of oil (J/m-s-K)	0.13
Thermal conductivity of gas (J/m-s-K)	0.04
Volumetric heat capacity of surrounding formation overburden and underburden (J/m^3-K)	66,465
Production time (yr)	16

*Viscosity *versus* temperature curve was obtained from *Barillas, 2005*.

Equations from mathematical modeling are embedded in STARS, which is a commercial software. Thus, it is not necessary to insert them in the input data file. Description of the process is done using specific keys from STARS library. Input parameters depend on available rock-fluid properties, as ones showed in **Table 1**. Because ideal models were used, they were not compared with real models.

2.2. Analyzed Operational Parameters

The minima (-1), intermediates (0) and maxima (+1) levels as well as nomenclature of analyzed operational parameters are shown in **Table 2**. A factorial planning of three levels was realized to study the effects of parameters interactions, resulting in 81 simulations.

Table 2. Range of analyzed operational parameters.

Parameter	Minimum (-1)	Intermediate (0)	Maximum (+1)
Distance between injector/producer wells (m)	65	100	135
Steam injection rate (m^3/day)	20	35	50
Solvent type	Pentane	Hexane	Heptane
Injected solvent volume/Injected steam volume (%)	5	10	15

Used maximum production rates correspond to the double of respective values of fluids (steam + solvent) injection rates. Oil recoveries (OR) were calculated dividing cumulative oil (CO) by original oil *in place* (OOIP) associated to each grid configuration (OR = CO/OOIP). Being a function of rock total volume, porous volume also varied, while all the others reservoir properties remained constant.

3. Results

Through the analysis of parameters interactions and taking oil recovery as objective function, an optimized model was achieved. **Table 3** shows the parameters from this model and others models used to emphasize the effect from injected solvent volume and to make a comparison with the case in which steam is injected without solvent.

Table 3. Operational parameters values for models 0%, 5%, 10% and 15% of injected solvent.

Parameter	0% Solvent	5% Solvent	10% Solvent	15% Solvent (Optimized Model)
Distance between injector/producer wells (m)	100	100	100	100
Steam injection rate (m³/day)	20	20	20	20
Solvent type	_	Heptane	Heptane	Heptane
Injected solvent volume/Injected steam volume (%)	0	5	10	15

3.1. Comparison among Primary Recovery and 0%, 5%, 10% and 15% Solvent Models
In **Fig. 2**, cumulative oil (without solvent) *versus* time are showed for primary recovery and models listed in **Table 3**. From second year on, curves with solvent models registered cumulative oil (without solvent) higher than the ones obtained by model 0% Solvent. This suggests that presence of solvent, for the analyzed amounts, accelerated the arrival of warm oil as far as producer well, promoting, thus, an earlier production with regard to model 0% Solvent. Moreover, it can be evidenced that an increase of injected solvent volume improves oil recovery (70,3% for model 5% Solvent, 71,1% for model 10% Solvent and 72,4% for model 15% Solvent).
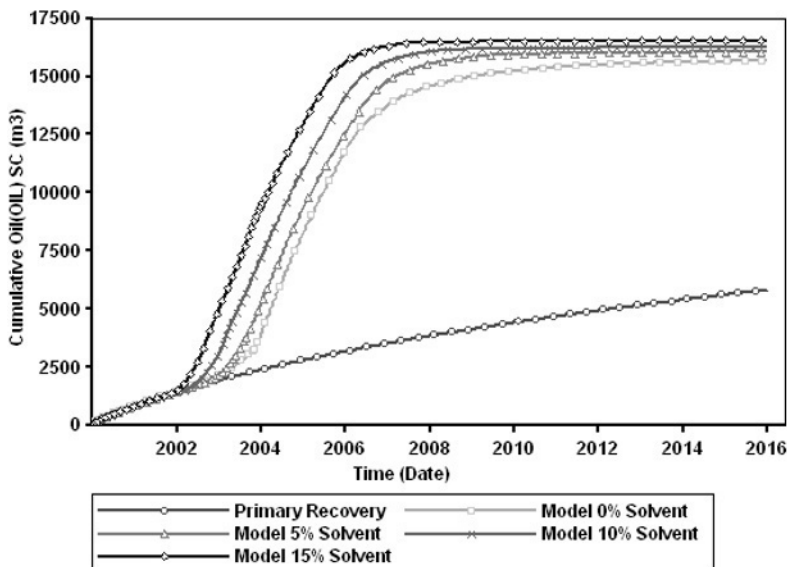
Fig. 2. Cumulative oil (without solvent) *versus* time. Primary recovery and models 0%, 5%, 10% and 15% Solvent.

Fig. 3 shows oil component (without solvent) rates *versus* time for these models. It can be observed that the higher injected solvent volume, the earlier production rates. This effect occurred mainly between second and fourth year, in which Optimized Model (15% Solvent) achieved 15.0 m^3/day, while model 0% Solvent achieved only 5.0 m^3/day at the same time. High rates obtained initially by Optimized Model contributed for a faster reservoir depletion, justifying the lower rates achieved by one in the following period.
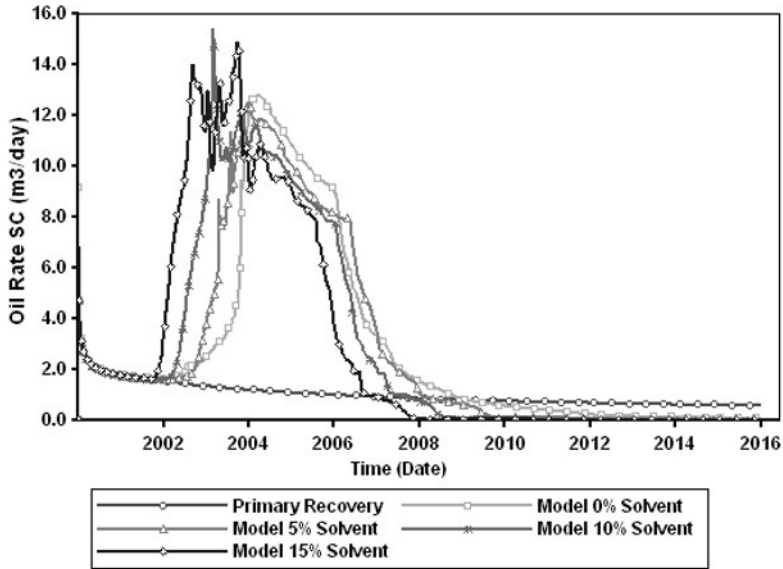
Fig. 3. Oil component (without solvent) rates *versus* time. Primary recovery and models 0%, 5%, 10% and 15% Solvent.

High initial productions achieved by models that use solvent have normally a significant impact on the operation economics, because early production suggests that fluids injection (steam and solvent) can be interrupted earlier. On environmental point of view, solvent injection can provide a reduction of energy and also a reduction in water consumptions for steam generation, having diminished Green House Gases (GHG) emissions. Also it is important to emphasize that the higher oil rates presented by these models can generate an earlier financial return and, by consequence, a project with a good economical viability (Galvão, 2008).

Injected and produced volumes of solvent *versus* time for models are registered in **Fig. 4**. Practically all injected solvent was produced together with oil from reservoir. To quantify the reuse of produced solvent is not in the scope of this work, however it is expected that produced fluid is lighter (high ºAPI) and has higher economic value (Galvão, 2008). Also it can be evidenced that the higher injected solvent volume, the earlier its recovery began.
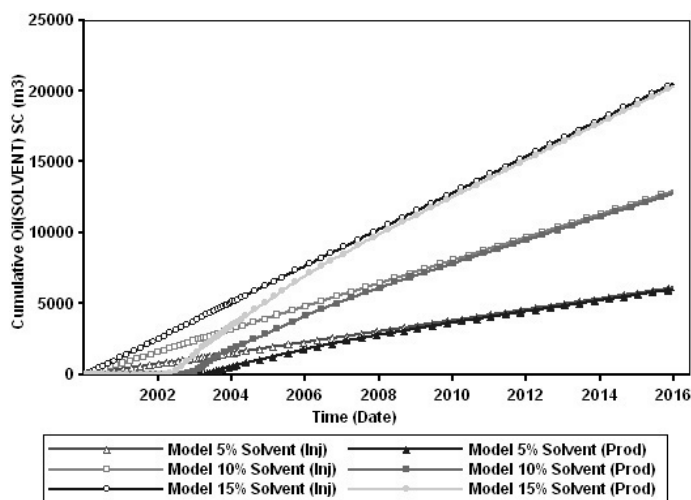
Fig. 4. Injected and produced volumes of solvent *versus* time. Models 5%, 10% and 15% Solvent.

4. Conclusions

• An increase of injected solvent volume improved the cumulative oil production with regard to model without solvent;

• Presence of solvent, for the analyzed amounts, accelerated the arrival of warm oil as far as producer well, promoting, thus, an earlier production with regard to model without solvent. This suggests that fluids injection (steam and solvent) can be interrupted earlier, reducing energy and water consumptions for steam generation, Green House Gases (GHG) emissions and generating an earlier financial return;

• Practically all injected solvent was produced together with oil from reservoir. It is expected that produced fluid is lighter (high ºAPI) and has higher economic value. Also it could be evidenced that the higher injected solvent volume, the earlier its recovery began.

Acknowledgements

References
NASR, T. N.; AYODELE, O. R. New hybrid steam-solvent processes for the recovery of heavy oil and bitumen. SPE 101717. Nov, 2006.

GALVÃO, E. R. V. P. Injeção de vapor e solvente como um método de recuperação avançada em reservatórios de óleo pesado. 2008. 106p. Dissertation (Master Degree in Petroleum Engineering) – Centro de Tecnologia, Programa de Pós-Graduação em Ciência e Engenharia de Petróleo, Universidade Federal do Rio Grande do Norte, Natal.

CMG, Computer Modelling Group Ltda. Guía para el usuario. Steam, Thermal and Advanced Reservoir Simulator - STARS. Versão 2007.11, Calgary-Alberta-Canada.

BARILLAS, J. L. M. Estudo do processo de drenagem gravitacional de óleo com injeção contínua de vapor em poços horizontais. 2005. 163p. Dissertation (Master Degree in Chemistry Engineering) – Centro de Tecnologia, Departamento de Engenharia Química, Programa de Pós-Graduação em Engenharia Química, Universidade Federal do Rio Grande do Norte, Natal.

Multi Objective Optimization Using Life Cycle Environmental Impact and Cost in the Operation of Utility Plants

Pablo Martínez, Ana María Eliceche

Departamento de Ingeniería Química,Universidad Nacional del Sur, PLAPIQUI-CONICET, Camino La Carrindanga km 7, 8000 Bahía Blanca, Argentina.

Abstract

Environmental and economic objective functions are used simultaneously to select the operating conditions of a steam and power plant. Different methodologies to solve multi objective optimization problems were implemented successfully. The life cycle potential environmental impact and the operating cost of the power plant are minimized simultaneously. A methodology is presented to estimate the potential environmental impacts during the most important life cycle stages associated with imported fuel and electricity in the utility plant. Mixed Integer Non Linear multi objectives problems are formulated and different strategies are implemented and successfully solved in GAMS.

Keywords: multi objective optimization, life cycle environmental impact, utility plant.

1. Introduction

The reduction of environmental impact of process operations is one of the imperative challenges to achieve sustainable development in process industry for the current century; also economical and social objectives are imperative issues for sustainability. Economic objectives have been used in process system engineering. However, environmental and economical objectives do not follow comparable behaviours. While the economic implication of a process is minimized, the associated environmental impact rises and an environmental friendly process is often cost intensive. This fact evidences the need of solving problems with more than one objective, that is, to find solutions that satisfy environmental and economic objectives at the same time. This solutions could be found formulating Multi objective Optimization Problems in the decision making process. The multi objective optimization problem could, in theory, be solved using similar methods as those employed in single objective optimization problems, converting the multiple objectives into a single objective. Multi-objective optimization applied to environmental and economic objectives has been treated by authors like Ciric and Huchette (1993) minimizing the amount of waste and the net profit of an ethylene glycol production plant. Dantus and High (1999) proposed a method to convert a bi objective optimization problem into a single objective optimization problem; the method proposed is a variation of the utopia point distance minimization, including discrete variables to select the type of reactor to be used in the methyl chloride superstructure plant design. Pistikopoulos and Hugo (2005) have treated multi objective optimization with environmental and economic objectives applied to supply chain network design and planning using the e-constraint method.

In the present work the operating conditions of a steam and power plant are selected to minimize life cycle environmental impact and operating cost simultaneously solving a multi objective optimization problem. The environmental objective is the life cycle environmental impact associated with solid wastes, gaseous and liquid emissions of a steam and power plant. In the life cycle context, the battery limits of the steam and power plant need to be extended in order to include emissions of imported natural gas and electricity generated by nuclear, hydroelectric and thermoelectric plants. The operating cost includes costs of imported electricity, natural gas feed, makeup water and water treatment. A Mixed integer non linear multi objective optimization problem is formulated and solved in GAMS (Brooke et al, 2003).

2. Environmental and Economic Objective Functions

2.1. Potential Environmental Impact Evaluation

The Potential Environmental Impact (PEI) function considered is a multi objective function itself, since nine environmental impact categories are considered: global warming, acidification, ozone depletion, photo oxidant formation, eutrophication, fresh water ecotoxicity, human toxicity, source depletion and the impact due to ionizing radiation. The Potential Environmental Impact is calculated using the Guinée et al. (2002) methodology. The contribution of the emission of a pollutant k to a given environmental impact category j is evaluated multiplying the pollutant k flow rate F_k emitted into the environment by a characterization factor γ_{kj} published by Guinée et al. (2002). This characterization factor represents the effect that chemical k has on the environmental impact category j. Hence, the Potential Environmental Impact, PEI, is calculated as follow:

$$\mathbf{PEI} = \sum_{j} \sum_{k} \boldsymbol{\alpha}_j \times \mathbf{F}_k \times \boldsymbol{\gamma}_{k,j} \tag{1}$$

Where α_j represent the weighting factors for each environmental impact category j. More information can be found in Eliceche et al. (2007). Eq.1 transforms the pollutants emissions flow rates into potential environmental impacts.

2.1.1. Utility Plant Environmental Impact

The emissions of the steam and power plant are evaluated from the modelling of the main processes formulated in GAMS. The emissions come mainly from the combustion in the boilers of a mixture of natural gas, F_{ng} and residual gas, F_{rg}. Liquid emissions of purge streams, F_p, in the boilers and cooling system are also considered. The pollutants emissions from the utility plant (UP) are calculated as follow:

$$\mathbf{F}_k^{\mathbf{UP}} = \mathbf{F}_{gn} \times \mathbf{e}_{k,ng} + \mathbf{F}_{rg} \times \mathbf{e}_{k,rg} + \mathbf{F}_p \times \mathbf{e}_{k,p} \tag{2}$$

Where $e_{k,ng}$ is the emission factor for pollutant k due to the combustion of natural gas, $e_{k,rg}$ is the corresponding emission factor for residual gas combustion and $e_{k,p}$ is the pollutant emission factor for liquid emissions. The emissions factors express the amount of pollutant k emitted by unit mass of natural gas, residual gas and liquid stream, respectively. The CO_2 emission due to natural gas and residual gas combustion are estimated stoichiometrically with the gas composition following the IPCC 2001

recommendations. Nearly 100 gaseous pollutants emissions are estimated from AP-42 report (EPA, 1998). Emissions from liquid discharges were estimated from the Electrical Power Research Institute report (2000).

2.1.2. Life Cycle Environmental Impact

Life cycle approach considers emissions during the entire life cycle of a product or service accounting by emissions from raw material extraction to waste disposal. In the case study presented in this work, the life cycle emissions are considered for the natural gas feedstock and the imported electricity needed to move some electrical motors in the superstructure of the steam and power plant. Pollutant flow rate for natural gas (NG) life cycle F_k^{NG} is calculated in the following equation:

$$F_k^{NG} = F_{ng} \times \sum_l e_k^l \qquad l = 1,...,l_{ng} \qquad (3)$$

Where e_k^l is the emission factor for pollutant k in the life cycle stage l, l_{ng} is the total number of life cycle stages considered for the natural gas fuel cycle: exploration, extraction and transportation stages. As the residual gas is produced in the ethylene plant, no life cycle stage has been considered for it.

For imported electricity (IE) life cycle, emissions have been assessed through the life cycle of different electricity generation plants. The electricity generation sector in Argentina has contributions from thermoelectric, hydroelectric and nuclear plants. Thermoelectric power generation consumes coal, oil and natural gas as fuels; nuclear power generation consumes natural uranium fuel. The estimation of pollutant emissions in the electric power generation includes the following life cycle stages: extraction and processing of raw materials, transport, refining (where it is applicable) and electricity generation itself:

$$F_k^{IE} = \sum_q \sum_{l_q} W_q \times e_{k,q}^{l_q} \qquad l = 1,...,l_{ie} \qquad (4)$$

Where W_q is the electricity imported and generated with technology q, l_q superscript accounts life cycle stage l in electricity generated by option q, finally $e_{k,q}^{l_q}$ is the corresponding emission factor of pollutant k in electricity generated with option q, for the life cycle stage l_q. The life cycle stages considered are: (i) exploration, extraction, refining and transport of natural gas, oil, coal and uranium consumed in thermoelectric and nuclear plants; (ii) submerged biomass decay in hydroelectric plants (iii) waste treatment and disposal for nuclear plants and (iv) transport in the construction stage of hydroelectric and nuclear plants.

The utility plant potential environmental impact, PEI^{UP} is calculated as follows:

$$PEI^{UP} = \sum_j \sum_k \alpha_j \times F_k^{UP} \times \gamma_{k,j} \qquad (5)$$

The component k life cycle emissions F_k^{LC} are estimated adding the component k emissions in the utility plant, life cycle of imported natural gas and electricity:

$$F_k^{LC} = F_k^{UP} + F_k^{NG} + F_k^{IE} \tag{6}$$

The life cycle potential environmental impact is evaluated as follows:

$$PEI^{LC} = \sum_j \sum_k \alpha_j \times F_k^{LC} \times \gamma_{k,j} \tag{7}$$

Global warming due to combustion emissions is the most relevant environmental category for steam and power plants and for fossil fuels electricity generation.

2.1.3. Economical Objective Function

The operating cost of the utility plant includes costs of imported electricity (IW), natural gas feed (NG), makeup water (MW) and water treatment (WT); where c_{ng}, c_q, c_{MW} and c_{WT} are the cost coefficients:

$$C = F_{ng} \times c_{ng} + \left(\sum_q W_q \right) \times c_W + F_{MW} \times c_{MW} + F_{WT} \times c_{WT} \tag{8}$$

A detailed mathematical model of the utility plant operation is presented in Eliceche et al (2007).

3. Multi objective Optimization with Environmental and Economic Objectives

The multi objective (MO) optimization is a system analysis approach to problems with conflictive objectives. A key factor of MO optimization is that rarely exist a single solution that simultaneously optimizes all the objectives. In its place, there is a set of solutions where one objective cannot be improved except at expense of another objective. This set of compromise solutions are generally referred as non-inferior or Pareto optimal solutions. A variety of strategies to solve multi objective optimization problems exist, that can be found in Alves et al (2007). The general approach consists in converting the multiple objectives into a single objective. Some of these methods are: weighted sum, utopia point distance minimization, e-constraint method and global criteria method. The general formulation of a bi objective optimization problem considering continuous and discrete variables follows:

$$\underset{x,y}{\text{Min }} Z = Z[PEI(x,y), C(x,y)]$$

$$\begin{aligned} \text{s.t}: \quad & h(x) = 0 \\ & g(x) + A(y) \leq 0 \\ & x^{LB} \leq x \leq x^{UB} \\ & x \in R^n \\ & y \in \{0,1\}^m \end{aligned} \qquad P1$$

Where x and y are the continuous and binary optimization variables, respectively. Superscripts U and L, indicates upper and lower bounds on vector x, respectively. The equality constraints $h(x) = 0$ are the system of non-linear algebraic equations that represent the steady state modelling of the process plant, including mass and energy balances; enthalpy and entropy prediction. The inequality constraints $g(x) + A(y) \leq 0$

represent logical constraints, minimum and maximum equipment capacities, operating and design constraints, etc. The A matrix includes linear relations between binary variables such as logical constraints.

Different strategies to solve multi objective optimization problems have been implemented successfully. The multi objective function Z in problem *P1* for the global method presented by Dantus and High (1999) follows, with the nomenclature presented in section 2:

$$\mathbf{Z} = \boldsymbol{\omega}_1 \times \left[\frac{\mathbf{PEI} - \mathbf{PEI}^*}{\mathbf{PEI}^{**} - \mathbf{PEI}^*} \right]^{\mathbf{p}} + \boldsymbol{\omega}_2 \times \left[\frac{\mathbf{C} - \mathbf{C}^*}{\mathbf{C}^{**} - \mathbf{C}^*} \right]^{\mathbf{p}} \qquad (9)$$

Where $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are weighting factors, these preference weights $\boldsymbol{\omega}_i$ are used to represent the relative importance of each objective. The decision-maker's preferences are also expressed in the compromise index $\mathbf{p}$ ($1 \leq \mathbf{p} \leq \infty$), which represents the decision-maker's concern with respect to the maximal deviation from the utopia point. As a result, the non-inferior solutions defined within the range $1 \leq \mathbf{p} \leq \infty$ correspond to the compromise set from which the decision maker still has to make the final choice to identify the best compromise solution (Dantus and High, 1999). The single asterisk indicates the minimum values of a given objective function solving a single objective optimization problem, while double asterisk indicates the alternative objective function value obtained. The objective functions used are life cycle potential environmental impact (PEI) given in Equation 6 and operating cost (C) given in Equation 8.

4. Numerical Results

A rigorous modelling of the utility plant is formulated in GAMS, including the power and steam demands of the ethylene plant (Eliceche et al, 2007). The continuous operating variables selected are temperature and pressure of the high, medium and low pressure steam headers and the deareator pressure. Binary operating variables are introduced to represent discrete decisions such as the selection of: (i) alternative pump drivers such as electrical motors and steam turbines and (ii) boilers which are on or off, and their auxiliary equipment such as feed pumps and air fans. Thus a multi objective Mixed Integer Nonlinear Programming problem is formulated and solved in GAMS.

Different strategies were implemented to solve the multi objective problem The solution point reported in Table 1 was obtained with the Dantus and High (1999) method and the following parameters for equation 9: $\boldsymbol{\omega}_1$=0.1, $\boldsymbol{\omega}_2$=0.9, $\mathbf{p}$=1. The following GAMS options were used: DICOPT as the outer approximation algorithm; CONOPT3 to solve the Non Linear Programming sub problem and CPLEX to solve the Mixed Integer Linear Programming sub problem.

Table 1. Multiobjective problem solution.

Objective Functions		Initial Point	Solution point	Reductions
PEI^{LC}	(PEI / h)	33627.33	29581.88	**12 %**
Cost	(U$A / h)	561.84	470.97	**16 %**

Significant reductions in the order of 12 % in the life cycle environmental impact and 16 % in the operating cost can be achieved selecting the operating conditions with the methodology proposed. Regarding the selection of pump's drivers, steam turbines are chosen rather than electrical motors, due to the fact that the environmental impact to

power generated ratio is smaller in the steam and power plant than in the generation of the imported electricity. This is due to the fact that natural gas is burned with residual gas from the demethanizer column. The residual gas is a Hydrogen rich stream, having higher combustion heat and lower combustion emissions than natural gas or any other fossil fuel. The operating cost is also cheaper with steam turbines than with electrical motors. The number of the boilers in operation is reduced from four to three, due to a proper selection of temperature and pressure of steam headers, mainly the high pressure steam header.

This is a process where improving process efficiency, environmental impact and cost are reduced simultaneously. They are not conflictive objectives. This is not the case if the environmental impact evaluation is reduced to the battery limits, where minimizing environmental impact and operating cost leads to different solutions mainly regarding the selection of alternative drivers.

5. Conclusions

A methodology has been presented to select the operating conditions minimizing simultaneously life cycle environmental impact and operating cost, solving a mixed integer nonlinear multi objective optimization problem. Imported natural gas and electricity life cycle environmental impacts have been estimated. Different strategies to solve multi objective optimization problems were implemented successfully. The ethylene steam and power plant analyzed, has a relevant contribution to combustion emissions, global warming, consumption of non renewable fossil fuels and operating cost. Thus, significant improvements in the plant operation can be achieved as shown in Table 1, with the strategy presented. This is a plant where improving process efficiency, environmental impact and cost are reduced simultaneously. It is also very important to extend the battery limits to include life cycle analysis, when environmental objectives are used to support a decision making process.

References

A. Brooke, D. Kendrick, A. Meeraus and R. Raman, 2003, GAMS development Corporation. A user guide.

A. Ciric and S. Huchette, 1993, Multiobjective Optimization Approach to Sensitivity Analysis: Waste Treatment Costs in Discrete process Synthesis and Optimization problems, Ind. Eng. Chem. Res., 32, 2636-2646.

A. Eliceche, S. Corvalan, P. Martinez, 2007, Environmental life cycle impact as a tool for process optimization of a utility plant, Comp.& Chem. Eng., 31, 648-656.

Electrical Power Research Institute, 2000, Aqueous discharges from steam electric power plants, data evaluation, Updated Report. http://www.epri.com

E. Pistikopoulos and A. Hugo, 2005, Environmentally conscious long-range planning and design of supply chain networks, Journal of Cleaner Production, 13, 1471-1491.

J. Guinée, R. Heijungs, G. Huppes , R. Kleijn, A. Koning, L. van Oers, A. Sleeswijk, S. Suh, H. Udo de Haes (eds.), 2002, Handbook on Life Cycle Assessment. Operational Guide to the ISO Standards. Kluwer Academic Publishers, Dordrecht.

IPCC. Climate Change 2001: the scientific basis. Intergovernmental Panel on Climate Change. Cambridge University.

M. Alves and J. Climaco, 2007, A review of interactive methods for multiobjective integer and mixed-integer programming, European J. of Operatinal Research, 180, 99-115.

M. Dantus, K. High, 1999, Evaluation of waste minimization alternatives under uncertainty: a multi objective optimisation approach, Comp. Chem. Eng. 23, 1493-1508.

U.S. Environmental Protection Agency,1998, AP-42. Compilation of air pollutant emission factors.

Modeling of Flowcharts of Permeation Through Membranes for Removal of CO2 of Natural Gas

Andressa Nakao[a], Ana P. F. Macedo[a], Betina M. Versiani[a], Ofélia de Q. F. Araújo[a], José L. de Medeiros[a]

[a] Departament of Chemical Engineering, Escola de Química da Universidade Federal do Rio de Janeiro. Sala E-201, Bloco E, Centro de Tecnologia, Ilha do Fundão, CEP: 21949-900 Rio de Janeiro, Brazil.

Abstract

In this present work, a simulator of permeation membranes is presented, the Permeation Membrane Simulator (PMS), which is based on rigorous engineering principles including sound thermodynamics, fluid mechanics, heat transfer and mass transport. The use of hollow-fiber membrane units is studied in the context of CO_2 removal from natural gas, which seems to be the preferred configuration in offshore plants. The feed stream is considered multi-component, with real chemical species, enabling reliable performance evaluations at different scenarios of pressure, temperature and composition of the gas to be treated by permeation. The resulting differential equations are solved along the characteristic contactor size: (i) steady state mass balances for the species present in each phase (L and V); (ii) momentum balances in each stationary phase; (iii) force balances in each phase. Major simplifications arise from adopting steady state and one-dimensional space over the membrane contactor.

The process arrangements are based on recent studies concerning subsalt oil finds in the offshore Santos Basin, Brazil. A process flowsheet is simulated, comprising compressors and membrane modules, corresponding to a scenario of 4MMm³/d and 12% mol CO_2. The process is composed of three permeation and compression stages, producing one stream of commercial grade natural gas and another stream rich in CO_2 for Carbon Capture & Geological Storage – CCGS - purposes. Considering the simulation results of the scenarios presented, the application of membrane technology in offshore facilities is proven technically feasible.

Keywords: Simulation, Natural Gas, CO_2, Hollow-Fiber.

1. Introduction

The recent intensification of Global Climate Change indicators highlights the need of reducing Greenhouse Gas (GHG) Emission. In this scenario, the main strategy considered by world leaders is synthesized by CCGS - Carbon Capture & Geological Storage; i.e. the capture of industrial emissions of CO_2 followed by underground sequestration into geological formations such as deep saline aquifers. The CCGS is presented as a Bridge-Technology in the sense that it is a temporary solution that should include CO_2 emissions giving subside to the World Economy of Energy, currently based on carbon fossil burning, while new sustainable energy sources become mature for widespread using. Brazil recent finds of natural gas (NG) in the giant subsalt reserves of the Santos Basin contains an expressive quantity of CO_2 (Formilgi, 2008), requiring

offshore treatment for CO$_2$ removal to: (i) minimize GHG emissions; (ii) increase the heating value of NG for use as fuel; (iii) reduce the potential for corrosion in metal pipelines to transport gas (Engelien, 2004); and (iv) to meet regulation standards for commercial NG (ANP, 2008).

Hence, NG purification is a contribution to CCGS strategies. Currently, .the main technologies used in industry are membrane modules and absorption in aqueous solutions of ethanolamines. For offshore platforms, membranes offer comparative advantages over other separation technologies – e.g. absorption and adsorption - such as: (i) simple operation requiring no heat exchange in either phase, (ii) free of chemical additions such as corrosive, toxic or flammable compounds, (iii) small foot-print due to compact, low weight units; and (iv) modular arrangements, which turns the process a flexible technology (Baker et al, 2008). These characteristics make the technology in permeation membranes more appropriate to the capture of CO$_2$ in offshore presalt conditions. However, it should be noticed that the main obstacle to the use of membranes is the immense process scale required in presalt NG purification. It is apparent that membrane systems lose selectivity when operating under high flow rates.

The selection of membrane configuration depends in the specifications set to the separation process. This study presents a process simulator developed as aid for performance evaluation of hollow-fiber membrane modules, the most used configuration for separation with gas permeation in membranes in offshore plants.

2. Model Formulation

The formulation of mathematical models capable of describing membrane separation processes is not trivial. The challenge is to develop models that contain the fundamental equations that describe the pertinent phenomena of this type of separation, with simple specifications, targeting potential industrial applications. The model premises are: (i) steady state; (ii) one independent variable – axial position z; (iii) rigorous 1-D compressible flow; (iv) heat transfer; (v) and permeation fluxes described via linear driving forces in terms of fugacities.

In this work, the retentate is represented by V and the permeate by L. Figure 1(a) shows the process configuration used and Figure 1(b) presents the membrane module used in the simulator, with variable labels employed in the model equations presented.

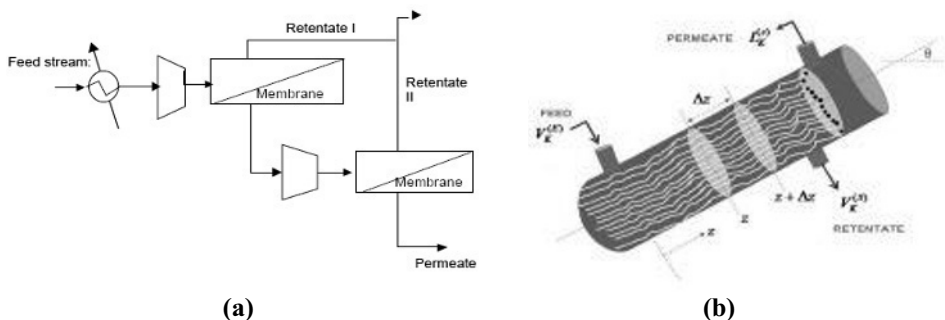| (a) | (b) |

Figure 1: Simulated Process. (a) membranes arrangement; (b) membrane module used in the simulator

The basic features of this model are: (i) independent variable: z; (ii) four dependent variables: pressure (P), temperature (θ), fluid inventory (I); and flow rates of species k (V_k, L_k); (iii) heat transfer taken into account from the distribution of external temperature, $?_E(x)$, and the distribution of heat transfer coefficient O(x); (iv) friction term via Churchill Equation for universal Darcy friction factor (*f*); (v) ideal gas properties according to Poling, Prausnitz and O'Connel (2001); (vi) viscosity of dense compressible fluid model proposed by Chung *et al.* (1988); (vii) rigorous thermodynamics via Peng-Robinson equation of state. Mass, momentum and energy balances are presented in Equations 1 through 7, derived in the hull's axial direction.

Mass Balance - Phase V:

$$\frac{\partial V_K}{\partial z} = - S\ a\ N_K \tag{1}$$

Mass Balance - Phase L:

$$\frac{\partial L_K}{\partial z} = S\ a\ N_K \tag{2}$$

Momentum Balance - Phase V:

$$\left\{1 - \Gamma_{P_V}\left(\frac{q_V}{\rho_V S_V}\right)^2\right\}\frac{\partial P_V}{\partial z} + \left\{-\Gamma_{T_V}\left(\frac{q_V}{\rho_V S_V}\right)^2\right\}\frac{\partial T_V}{\partial z} = \frac{S\ a\ q}{S_V}\left(\frac{q_V}{\rho_V S_V}\right) - \rho_V g\ sen(\theta) - \frac{\Psi_V(\pi D + S\ a)}{S_V} \tag{3}$$

Momentum Balance - Phase L:

$$\left\{1 - \Gamma_{P_L}\left(\frac{q_L}{\rho_L S_L}\right)^2\right\}\frac{\partial P_L}{\partial z} + \left\{-\Gamma_{T_L}\left(\frac{q_L}{\rho_L S_L}\right)^2\right\}\frac{\partial T_L}{\partial z} = \frac{S\ a\ q}{S_L}\left(\frac{q_L}{\rho_L S_L}\right) - \rho_L g\ sen(\theta) - \frac{\Psi_L S\ a}{S_L} \tag{4}$$

where

$$\Gamma_{P_V} = \left(\frac{\partial \rho_V}{\partial P}\right)_{T,\underline{N}}, \Gamma_{P_L} = \left(\frac{\partial \rho_L}{\partial P}\right)_{T,\underline{N}}, \Gamma_{T_V} = \left(\frac{\partial \rho_V}{\partial T}\right)_{P,\underline{N}}, \Gamma_{T_L} = \left(\frac{\partial \rho_L}{\partial T}\right)_{P,N} \tag{5}$$

Energy Balance - Phase V:

$$\left\{1 - \Gamma_{P_V}\left(\frac{q_V}{\rho_V S_V}\right)^2 + \frac{T_V \Gamma_{T_V}}{\rho_V}\right\}\frac{\partial P_V}{\partial z} + \left\{\frac{\rho \overline{C}_P^V}{M_V} - \Gamma_{T_V}\left(\frac{q_V}{\rho_V S_V}\right)^2\right\}\frac{\partial T_V}{\partial z} = -\rho_V g\ sen(\theta) + \frac{S\ a\ q}{S_V}\left(\frac{q_V}{\rho_V S_V}\right) +$$
$$\frac{\Omega_E \pi D \rho_V}{q_V}(T_E - T_V) + \frac{\Omega\ S\ a\ \rho_V}{q_V}(T_V - T_L) \tag{6}$$

Energy Balance - Phase L:

$$\left\{1 - \Gamma_{P_L}\left(\frac{q_L}{\rho_L S_L}\right)^2 + \frac{T_L \Gamma_{T_L}}{\rho_L}\right\}\frac{\partial P_L}{\partial z} + \left\{\frac{\rho \overline{C}_P^L}{M_L} - \Gamma_{T_L}\left(\frac{q_L}{\rho_L S_L}\right)^2\right\}\frac{\partial T_L}{\partial z} = -\rho_L g\ sen(\theta) + \frac{S\ a\ q}{S_L}\left(\frac{q_L}{\rho_L S_L}\right) +$$
$$\frac{\Omega\ S\ a\ \rho_L}{q_L}(T_V - T_L) + \frac{S\ a\ \rho_L}{q_L}\sum_K N_K\left(E_K^V - E_K^L\right) \tag{7}$$

3. Case Study

The flowsheet GNBRPS presented in Figure 3 is based on studies from Santo Basin, Brazil (Formigli, 2008), and defines a process fed with current GNBRPS with average content of CO_2 (12% mol), pressure of 60 bar, temperature 25°C and flow rate of 4 MMm³/d, comprised of 3 permeation modules with intermediary compression operations. The compression operates in 3 stages with efficiency of 75% and intercoolers to 35 °C. Process specification is to perform the permeation at low pressure in module 1 (P = 1bar) before feeding the module 2.
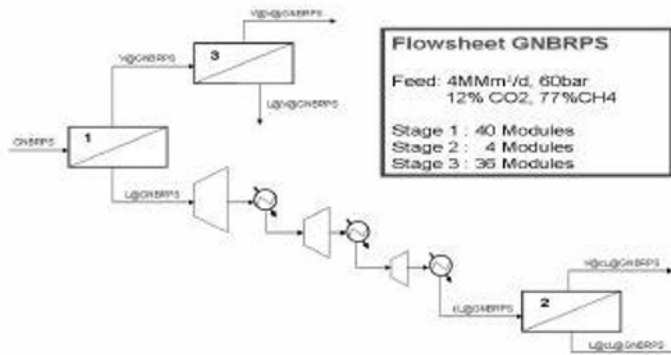
Figure 3: Flowsheet GNBRPS

Simulation Results for First Stage: Figures 4(a) and 4(b), respectively, show composition profiles (mol %) of CO_2 and CH_4 in both phases V and L of the process. Figure 4(c) illustrates the profile (%) of CO_2 recovery in phase L (permeate 1). In phase V, there is a considerable removal of CO_2, from 12% of CO_2 in feed, to 6% CO_2 content. According to Brazilian regulation agency (ANP, 2008), commercial NG is required to show a maximum of 3% mol CO_2, and phase V still does not fit within the specifications for marketing. This product, V@GNBRPS, is hence submitted to a second permeation. In phase L, CO_2 represents 67% of the permeate, and a recovery of 55% of CO_2 in phase L is achieved, requiring a further permeation module (corresponding to Stage 2) to increase the content of CO_2.

| (a) | (b) | (c) |
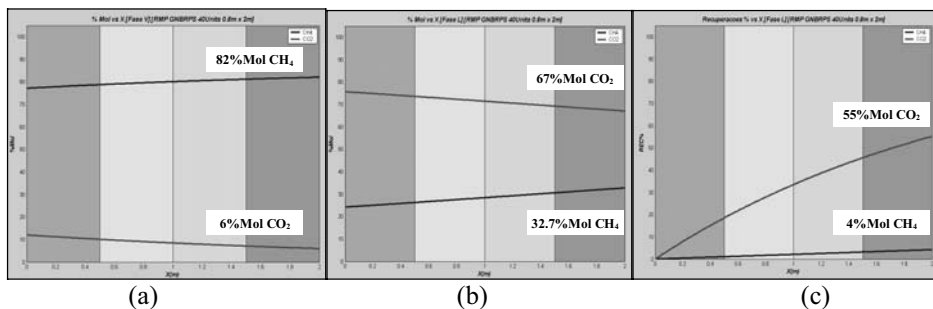
Figure 4: Stage 1 - (a) Profile (mol %) of phase V; (b) Profile (mol %) of phase L; (c) Profile (mol %) of CO_2 recovery in phase L

Simulation Results for Second Stage: As the content of CO_2 in phase L is not high enough to meet specification for injection into reservoirs, a second permeation stage is required. Furthermore, the pressure of permeate 1 is 1 bar, demanding recompression to 60 bar, by means of a train of compression as a single stage operating from 1 to 60 bar. The profiles displayed in Figure 5 represent the mol% of the two gases studied (CO_2 and CH_4) in both phases of the process (Figures 5(a) and (b), respectively) and the profile of recovered CO_2 in phase L (permeate) (Figure 5(c)). In phase V, - stream V@cLGNBRPS, a good removal of CO_2 is obtained: from 67% of CO_2 in the feed to 38%. This stream unfortunately is not specified for use in any of the key objectives (commercial GN or CO_2 for injection), although it can be used as burning gas. For stage L, CO_2 reached nearly 96.3% of the composition of the permeate, a level acceptable to

be used for injection in reservoirs. Therefore, it reached one of the objectives of this work, which would be to obtain a stream of CO_2 for injection. There was a recovery of almost 72.2% of CO_2 in phase L.
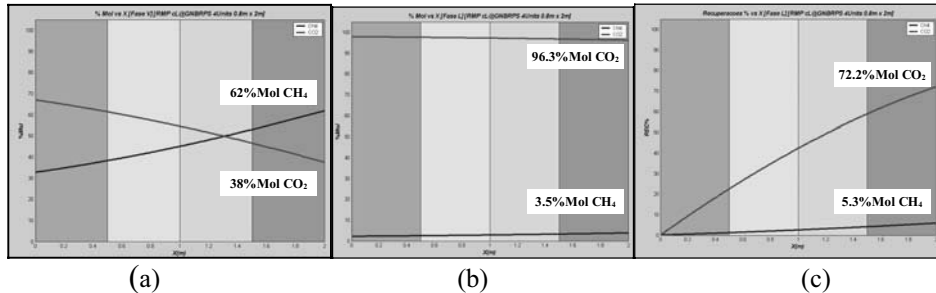

Figure 5: Stage 2 - (a) Profile (mol %) of phase V; (b) Profile (mol %) of phase L; (c) Profile (mol %) of CO_2 recovery in phase L

Simulation Results for Third Stage: This stage is fed with the retentate of Stage 1, which is at 59.62 bar, therefore not requiring further compression. This stream has 6% CO_2 and 82% of CH_4, equivalent to 90% of the original feed GNBRPS, justifying the use of 38 permeation modules in Stage 3. Figures 6(a) and (b) present the profiles of of CO_2 and CH_4 (mol%) in phases V and L, respectively. The profile of CO_2 recovery in phase L is shown in Figure 6(c). In phase V, stream V@V@GNBRPS shows a good removal of CO_2, based on a 6% CO_2 in the feed, slightly inferior to 3%, hence meeting the second process objective of this work - obtaining a specified stream to NG marketing. In phase L, CO_2 reached 46.5% of the composition of permeate and there was a recovery of nearly 51.5% of CO_2. This stream can still be used for burning or to turbines.
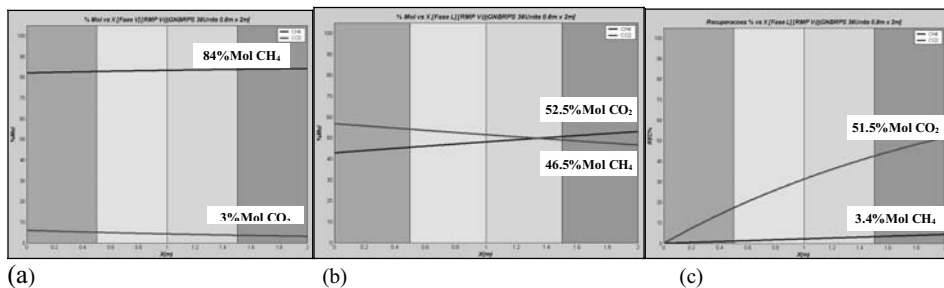
Figure 6: Stage 3 - (a) Profile (mol %) of phase V; (b) Profile (mol %) of phase L; (c) Profile (mol %) of CO_2 recovery in phase L

Table 1 2 summarize the results of GNBRPS simulated in PMS.

4. Conclusion

This work was presented and simulated a stream for purification by permeation membranes in a natural gas compatible with the scenario of the Pole Pre-Salt of the Santos Basin. Membrane modules of cellulose acetate in hollow-fiber were considered. The study found the technical feasibility of the use of membranes for CO_2 separation from it being a typical load of natural gas. This feasible technique is to obtain a stream

of treated natural gas suitable for marketing and a main stream of residual CO_2 suitable for geological injection.

Table 1: Stream specifications

Stream	MMm^3/d	P (bar)	T (^{o}C)	CO$_2$ (%)	CH$_4$ (%)
GNBRPS	4	60	25	12	77
V@GNBRPS	3.6	59.6	20	6	82
L@GNBRPS	0.4	1	19	67	32.7
cL@GNBRPS	0.4	60	35	67	32.7
V@cL@GNBRPS	0.2	59.7	4.5	38	62
L@cL@GNBRPS	0.2	1	2	96.2	3.5
V@V@GNBRPS	3.37	59.2	17.4	3	84
L@V@GNBRPS	0.23	1	17	46.5	52.5

Table 2: Simulation results

	MMm^3/d	CO$_2$ (%)	CH$_4$ (%)	RECOVERED CO$_2$ (%)	RECOVERED CH$_4$ (%)
Commercial NG	3.37	3	84	22%	91.9%
CO$_2$ for Injection	0.2	96.2	3.58	40%	0.2%
NG for Burning	0.2 + 0.23	38/46	62/52.5	38%	7.9%

Considering the simulation results of the presented case study, this work demonstrates the technical feasibility of applying membrane technology in offshore installations. In addition to the listed advantages of membrane technologies over traditional absorption columns, specified levels of CO_2 and CH_4 are herein proven to be achievable though the use of MPS, meeting process specifications and complying with environmental regulations.

References

Agencia Nacional de Petróleo e Biocombustível. Resolução 16, DOU 17.06.2008

R. W. Baker, K. Lokhandwala,2008. Natural gas processing with membranes: An Overview. Ind. Eng. Chem. Res., 41, p. 2109 – 2121

H. K. Engelien, 2004 Process Integrated Membrane Separation - with Application to the Removal of CO2 from Natural Gas. Department of Chemical Engineering, NTNU.

J. Formigli, 2007. Pre-Salt Reservoirs Offshore Brazil: Perspectives and Challenges

1881

Optimal Economic Decision Making for Gas-to-Liquid Product Selection Considering Competition in Market Dynamics

Chul-Jin Lee and Chonghun Han

School of Chemical and Biological Engineering, Seoul National University, San 56-1, Shillim-dong, Kwanak-gu, Seoul, 151-742, Korea

Abstract

For a chemical process industry, an invested capital cost is too large and when a decision making that which product should be produced is once determined, it is almost impossible to be modified or repeated. Also, since the huge uncertainty exists in the whole market such as recession today, it is so important to judge an economic decision making for selection of which target product would be the most profitable in the future. For the sake of a robust decision, we need to compare several points of view including profitability, reliability, risk, and so on. In this paper, we will try to get an optimal economic decision making for Gas-to-Liquid product selection considering market dynamics. The targets are three GTL products; FT-diesel, DME, and Methanol. At first we could design the three GTL processes using a process simulator. Then, through the constructed models, comparative economic analysis among them was conducted using an economic index. For a robust analysis we can apply the market dynamics related with demand expansion and shrinkage as well as a competition within players. Finally, we can get an optimal solution for selecting GTL product considering market variation and compare the robustness for the decision.

Keywords: decision making, process design, economic analysis, Gas-to-Liquid, Market dynamics

1. Introduction

There have been several economic analyses for GTL processes consuming natural gas as feedstock. Han et al. (2002) assessed an economic evaluation for methanol synthesis process and Zhou et al. (2007) studied the economic and environmental influences of DME process [1], [2]. In general, however, many people concentrated on individual analysis for a single GTL product, so there were few studies considering a comparative economic analysis to select the most profitable product using natural gas as raw material. Along the way Gradassi et al. (1995) compared the economic analysis of several GTL products, and Morita (2001) concentrated on three profitable GTL products (DME, FT-diesel, Methanol) used for transportation fuels. However, these studies did not cover the characteristics of process industry such as the fact that the profitability of each product is largely dependent on the raw material cost [3], [4]. Lee et al. (2009) focused on the influence of natural gas cost variation as raw material on the profitability of GTL products and found that the order of profitability of GTL products could be reversed according to the raw material cost variation [5]. However, the profitability of chemical products is largely influenced by consumer market. We need to consider the market dynamics of GTL to get an optimal economic decision for selecting GTL product.

In this paper, we modeled three GTL processes producing DME, FT-diesel, and Methanol using ASPEN PLUS^{TM} and performed an economic assessment of each of them. Furthermore considering the market variations of each product, we analyzed the profitability of them using net present value (NPV) as an economic index. As a result, the priority of profitability was reversed according to the market variations in the future.

2. Modeling and Simulation

The GTL process is largely divided into two parts. At first we could produce the synthesis gas which is a gaseous mixture of hydrogen and carbon monoxide from natural gas as raw material, which is called to reforming process. Then, we can synthesize the target chemical product such as FT-diesel, Methanol, and DME. Both stages of reaction go through in high temperature (~1000℃) and pressure (~3000kPa), the Peng-Robinson EOS was used as a thermodynamic package model. The detailed models and results for three GTL processes were described by Lee et al. [5].
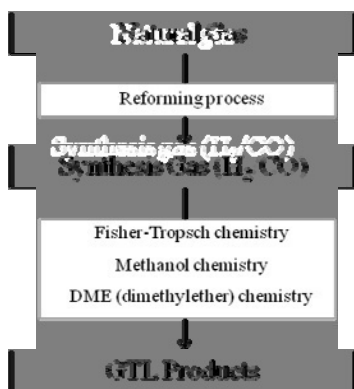
Figure 1. Schematic diagram of the Gas-to-Liquid process

3. Economic Evaluation

3.1. Previous economic analysis summary
Through the constructed process model, we can calculate a required capital and operating cost. The evaluation result of each product is shown in Table 1. The detail of each is presented by Lee et al. [5].

As the payout time means the required time to recover an invested capital, the case of the lower payout time means the more profitable one. Thus, producing FT-diesel is the most profitable among three GTL products. In this case, we assumed that all of the produced products are sold in the market. However, there can be a competition between the producers in market or the whole market size would be expanded or shrink according to the boom days or recession. To obtain an optimal decision making to select a prosperous GTL product, we need to consider a market competition and variation over time series.

Table 1. Summary of an economic analysis for three GTL products.

	unit	FT-diesel	MeOH	DME
revenue	$x10^6$/yr	$1,038	$844	$767
capital cost	$x10^6$	$2,374	$2,254	$1,289
utility cost	$x10^6$/yr	$8	$18	$11
NG cost	$x10^6$/yr	$606	$606	$606
transportation cost	$x10^6$/yr	$21	$49	$17
payout time	yr	5.91	13.24	9.76

3.2. Net Present Value

For an economic index in this paper, we used the Net Present Value (NPV) to consider a variation of cash flow to the investment period [6]. The value of minimum expected rate of return, r, was set to 5.7%, by taking account of the yield of treasury bonds over 10 years in South Korea [7].

$$NPV = \sum_{i=1}^{N} \frac{CF_i}{(1+r)^i} - I \tag{1}$$

4. Market Competition

Considering the market dynamics, we can assume the competition in each GTL market. For instance, the stronger player in GTL market will occupy more market share annually than the weaker one. Thus, the sales quantity of each would be varied to the expansion rate of market occupation of each, which can make change the revenue of them.

4.1. Assumption

(1) There are three players(P1, P2, P3) for FT-diesel, MeOH, and DME markets individually.

(2) The whole market size is fixed for each product during investment period.

$$Q_A = Q_{A1} + Q_{A2} + Q_{A3} \tag{2}$$
$$Q_B = Q_{B1} + Q_{B2} + Q_{B3} \tag{3}$$
$$Q_C = Q_{C1} + Q_{C2} + Q_{C3} \tag{4}$$

(3) Among the three players in each product, there exists a competition which can make vary the market portion of them to 10%, 0%, and -10% annually.

$$\frac{dQ_{Aj}}{dt} = \alpha_i, \text{ i = 1, 2, 3 for each player in FT-diesel market} \tag{5}$$

$$\frac{dQ_{Bj}}{dt} = \beta_i, \text{ i = 1, 2, 3 for each player in MeOH market} \tag{6}$$

$$\frac{dQ_{Cj}}{dt} = \gamma_i, \text{ i = 1, 2, 3 for each player in DME market} \tag{7}$$

4.2. Analysis results in considering competition among players in each market

The trends of NPV result for each GTL product are analogous to the previous economic result, which used the payout time as an economic index shown in Table 1. When the market of FT-diesel annually expands to 10%, the value of NPV of FT-diesel is slightly larger than DME case as depicted in Figure 2 and 4 ($\alpha1$ and $\gamma1$ respectively). However, when the market portion shrinks to -10% every year, which means the falling on competitive power or recession of the whole market, the difference of NPV is quite increased that a negative NPV of FT-diesel is almost double comparing DME value, which indicates the possible loss of FT-diesel production would be larger than DME case.
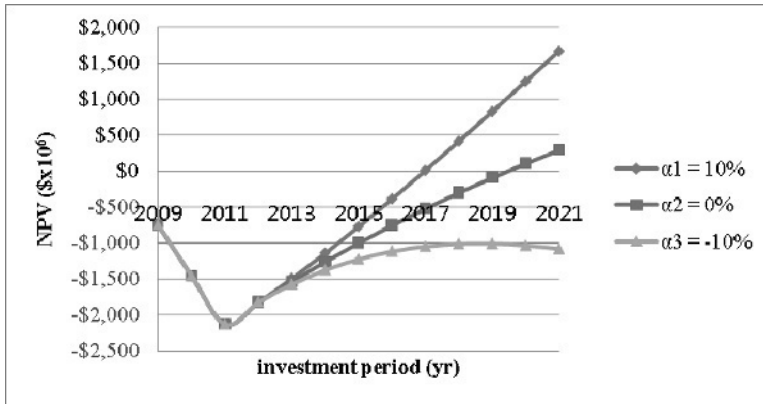

Figure 2. NPV variation of FT-diesel production over the service life for annual expansion rate of market share for FT-diesel player 1, 2, 3.
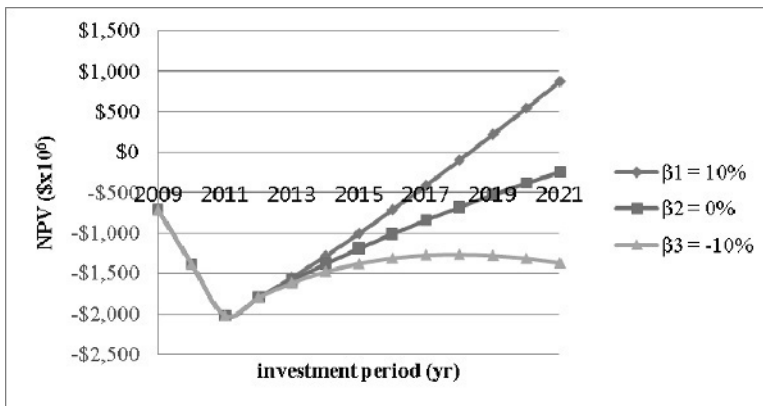
Figure 3. NPV variation of MeOH production over the service life for annual expansion rate of market share for MeOH player 1, 2, 3.
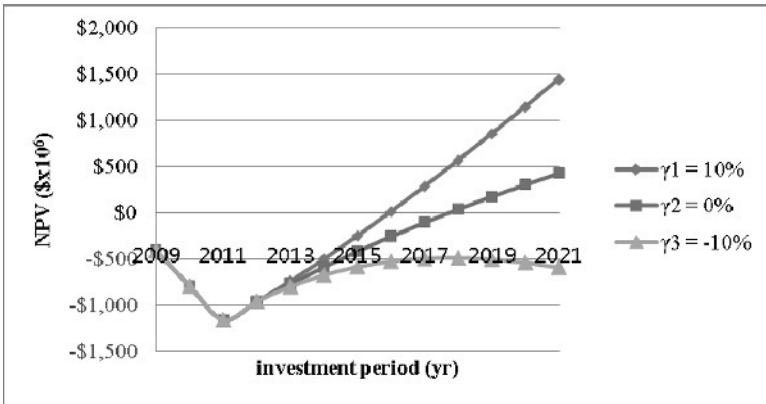
Figure 4. NPV variation of DME production over the service life for annual expansion rate of market share for DME player 1, 2, 3.
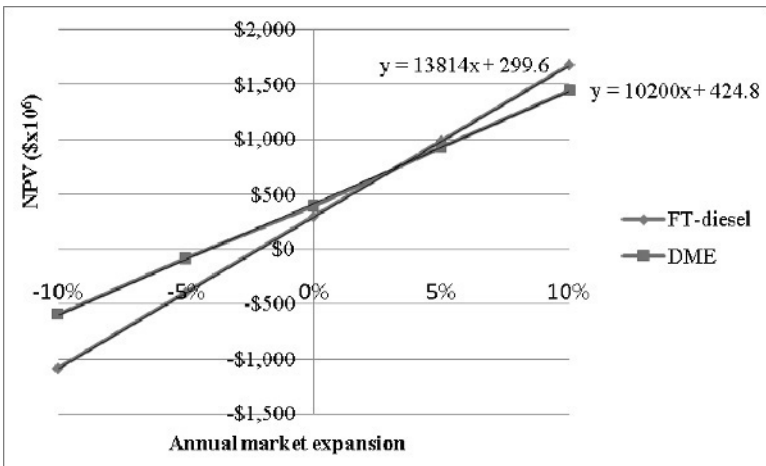
Figure 5. NPV comparison between FT-diesel and DME according to the annual market expansion.

Thus far, we could analyze the economic feasibility among three GTL products considering market competition among players. Then, we can guess how the profitability can be changed in taking account of the variations of whole market size. When there may be prosperous condition or recession in the whole market, which product would be the most robust decision? We can get the answer in Figure 5. Because the MeOH production always gets the lowest profitability, we can concisely focus on economic comparison between FT-diesel and DME to select an optimal product.

When we assume that both markets of FT-diesel and DME would be influenced by the same amount for expansion or shrinkage, we can get the correlation function between NPV and annual market expansion rate. The intersection position for both NPV graphs points out an expected market growth rate balancing the profitability of them. When the both markets grow more than 3.6% per annum, producing FT-diesel can get more profit. On the other hand, it is the better way to choose DME below 3.6% per annum, and further, it can reduce the expected loss at the lower position than the break-even point at zero value of NPV.

5. Conclusion

In this article, we investigated the optimal economic decision making for Gas-to-Liquid product selection when a competition between players in each market exists. Furthermore, when we expect the expansions or shrinkage of the whole market size, we can compare the profitability among GTL products. When the market expands over 3.6% per annum, it is better to choose the FT-diesel. On the contrary, DME production is more profitable below 3.6% per annum. Suggested methodology including market dynamics would be especially useful to decision makers when the risk and uncertainty of the market environment considerably matters.

Nomenclature

d_i = discount factors
CF_i = after taxes cash flow of period
I = the capital investment
r = minimum expected rate of return
Q_A = production quantity for FT-diesel
Q_B= production quantity for MeOH
Q_C= production quantity for DME
Q_{Ai} = production quantity of player i for FT-diesel
Q_{Bi} = production quantity of player i for MeOH
Q_{Ci} = production quantity of player i for DME
α_i = annual increase rate of market portion of player i for FT-diesel
β_i = annual increase rate of market portion of player i for MeOH
γ_i = annual increase rate of market portion of player i for DME

References

[1] Han, S. R.; Chung, Y.; Joo, O.; Jung, 2002, K. Economics of methanol synthesis process. Theor. Appl. Chem. Eng., 8, 3113-3116.
[2] Zhou, L.; Hu, S.; Li, Y.; Zhou, Q., 2008, Study on co-feed and co-production system based on coal and natural gas for producing DME and electricity. Chem. Eng. J., 136, 31-40.
[3] Gradassi, M. J.; Green, N. W., 1995, Economics of natural gas conversion processes. Fuel Process. Technol., 42, 65-83.
[4] Morita, Y., 2001, Marketability of GTL from Natural Gas; The Institute of Energy Economics: Japan.
[5] Lee, C. –J.; Lim, Y.; Kim, H. S.; Han, C., 2009, Optimal Gas-to-Liquid Product Selection from Natural Gas under Uncertain Price Scenarios. Ind. Eng. Chem. Res., vol. 48, no. 2., pp. 794-800.
[6] Miguel, Bagajewicz, 2008, On the Use of Net Present Value in Investment Capacity Planning Models. Ind. Eng. Chem. Res., vol. 47, no. 23., pp. 9413-9416.
[7] Lim, S. –R.; Park, D.; Lee, D. S.; Park, J. M., 2006. Econmic evaluation of a Water Network System through the Net Present Value Method Based on Cost and Benefit Estimation. Ind. Eng. Chem. Res., vol. 45, pp. 7710-7718.

1887

Bioprocess Systems Engineering Applied to the Production of Protein Hydrolysates in a Multipurpose Plant

Gilson A. Pinto,[a] Roberto C. Giordano[b]

[a]*Chemtech – a Siemens Company, Av. Ermano Marchetti, 1435, 05038-001, São Paulo – SP, Brasil*
[b]*Department of Chemical Engineering, Federal University of São Carlos, Rod. Washington Luis, Km. 235, São Carlos – SP, P.O.Box 676, Brasil*

Abstract – The continuous advances in process computing have provided in recent years several sophisticated tools for process analysis and simulation. The effective use of these tools demands the capability to interface with a pre-existent process system hierarchy. Within this perspective, an important issue is to provide meaningful and useful simulation and optimization applications for complex systems that require integration with data-intensive experimentation. This study proposes an integrated environment, using Internet as development platform, for simulation, monitoring, control and optimization of a cheese whey refinery, employing immobilized and stabilized enzymes as catalysts. The multipurpose process described here, the *cheese whey biorefinery*, provides, besides lactose, whey protein concentrates and hydrolysates that can be applied in food and pharmaceutical formulae. Under these circumstances, it is possible to add value to this significant by-product of the dairy industry, avoiding its disposal *in natura* (what is mostly done by small cheese manufacturers). In parallel, relevant aspects of the enzymatic reactions within the cheese whey biorefinery were investigated. The lumping of substrate molecules in pseudo-components, using a hybrid phenomenological-neural approach for description of the enzymatic de-polymerization kinetics, is the suggestion to follow the complex reaction dynamics. During the estimative of model parameters and stationary state simulation/optimization of the multipurpose plant, non-linear search algorithms were implemented and evaluated (Successive Quadratic Programming, Simulated Annealing, Differential Evolution and Particle Swarm Optimization). The work, thus, is concerned with two of the main dilemmas/challenges of the present millennium: reduction of environmental problems related to the disposal of agro-industrial residues (i.e., cheese whey) and development of processes for food production from alternative sources (whey refinery). All implemented algorithms for remote process monitoring and control were validated with data from laboratory-scale assays. The validation of the hydrolytic kinetic models followed the same procedure. According to the achieved results, is possible to conclude that the robustness of the integrated computational environment was demonstrated. The prediction capability of the approaches employed for description of the proteolytic enzymatic reactions was verified, as well.

Keywords: cheese whey biorefinery, enzymatic kinetic modeling, nonlinear optimization, Internet based application.

1. Introduction

The concept of a "biorefinery" is usually tied to the processing of lignocellulosic biomass to produce fuels and chemicals. Nevertheless, if we have in mind an oil refinery as our paradigm, any plant that has a pool of macro- and micromolecules for feedstock, providing a set of different products (usually, pools of molecules), using bioreactors (either enzymatic or biochemical – "fermenters") for the "refining" of the raw materials, any of these plants may be named a biorefinary as well. It is within this broader concept that we report in this work the application of (bio)process systems engineering tools to what we call a "cheese whey biorefinery".

A simplified block diagram of the whey biorefinery, as proposed in Pinto *et al.*, 2009, is shown in Fig. 1. In this work, we detail the integration of algorithms for plant-wide optimization, bioprocess remote monitoring and control (via internet), and hybrid models of complex de-polymerization enzymatic reactions into a user-friendly environment. The originality of the concept is putting together a static simulator (with design, optimization and cost analysis tools) and an applicative for remote monitoring and control of the bioprocess.
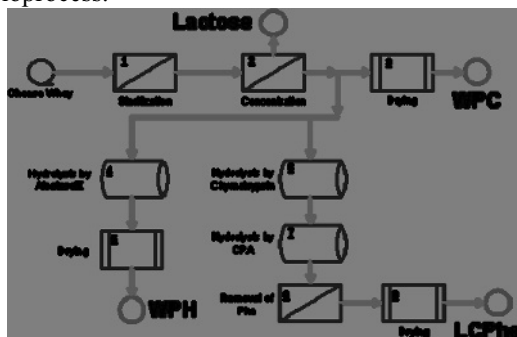

Fig. 1 – Cheese whey biorefinery block diagram.

The enzymatic customized hydrolysis of cheese whey may provide a collection of value-added products: Advances in bioprocess technologies have enhanced the use of cheese whey for commercial applications such as: edible films, whey protein concentrates, functional foods, athletes' supplements, clinical diets and pharmaceuticals (Clemente, 2000). In this way, a strong pollutant disposal of the dairy industry could be transformed into commercial products. Indeed, whey BOD is 35000 mgO$_2$.L$_{-1}$, approximately 10^2 times greater than domestic sewage's BOD; Siso, 1996.

In Fig. 1, raw cheese whey is micro- and ultra-filtered, and lactose is separated, following to further enzymatic hydrolysis (not shown). A whey protein concentrate is a possible first product of the biorefinery, after spray-drying (WPC). Alternatively, WPC may go through a series of controlled proteolyses with immobilized enzymes. Supplements for parenteral feeding come from the hydrolysis with alcalase®, providing a pool of small peptides with lower osmolality than a mixture of amino acids – WPH, Whey Protein Hydrolysates (Pinto et al., 2007). A different product, LCPhe (Low Contents of Phenylalanine) can be produced after two sequential enzymatic hydrolyses with chymotrypsin and carboxypeptidase A, with subsequent separation of amino acids and remaining lactose (Cabrera-Padilla et al., 2009; Galvão et al., 2009). This product can be used for feeding phenylketonuria patients.

The computational applicative previously mentioned has tools for the preliminary design, simulation, remote monitoring and control. The simulation module of the web application was developed to provide stationary mass balances, preliminary design and cost analysis of the main equipments. Economic analysis of the process is another

feature of the application, which includes an optimization module for the multipurpose plant. In the integrated Internet-based environment here presented, on-line monitoring and control modules were also put forth. The proteolytic steps of the biorefinery were investigated within this perspective, since real-time inference of the product composition is crucial for the fine-tuning the control of the enzymatic reactors. Given that mechanistic kinetic models of the enzymatic reactions would be very complex, hybrid neural-network models were employed. The prediction capability of the models and their implementation within the web application for real-time monitoring of the reactors are presented here.

2. Software Development and Implementation

2.1. Non-linear Search Algorithms

The approach used here was to implement the stochastic methods described below, in Matlab 6.5 (MathWorks, Inc.), coupled to a set of heuristic corrections for each implementation. The main idea was to use these techniques as tools for a real optimization problem, rather than concentrating a major effort in tuning the algorithms.

The initial temperature, T^0, of simulated annealing (SA) was selected according to Kirkpatrick et al., 1983, computing a temperature such that the initial acceptance ratio is approximately equal to a given value λ_0 (here, $\lambda_0 = 0.8$). First, a large value for T^0 is chosen. Then, a number of transitions using this temperature are performed. The ratio of accepted transitions is compared with λ_0. If it is less than the acceptance ratio, then T^0 is multiplied by 2. The procedure continues until the observed acceptance ratio exceeds λ_0.

The cooling scheduling for each iteration k was $T^{k+1} = 0.98 T^k$. The neighborhood structure was constructed by simple random values constrained by the lower and upper bounds of parameters. The implementation of particle swarm optimization (PSO) followed the pseudo-code described by Kennedy and Eberhart, 1995. Introduced by Shi and Eberhart, 1998, the inertial weight w – not present in the original form – was also used. According to the authors, the parameter w can assure the PSO convergence provided it decreases linearly at each iteration k, according to $w^k = 0.9 - 0.5(k / it_{max})$. The implemented strategy for Differential Evolution (DE) was DE/rand/1/binomial, as described by Storn and Price, 1997. The algorithm *fmincon* from Matlab 6.5 was used for SQP (sequential quadratic programming).

The optimization of stationary states of biorefinery – each whey flow rate defines a problem (stream *1* in Fig. 1) – is solved with the definition of two auxiliary variables, showed in the process flowsheet described in Fig. 1: p_{WPC} and p_{WPH}, respectively, the percentage of stream *7a* that will be directed to WPC production and the percentage of stream *15a* that will be used for WPH production. The net present value of the project, *NPV*, is implicitly a function of p_{WPC} and p_{WPH}, since these variables define stationary process flow rates and, consequently, the costs and yields involved in the biorefinery project (equipment dimensions, capital costs, etc.). Therefore, the nonlinear programming problem can be stated as:

$$\underset{p_{WPC}, p_{WPH}, \tau_{R-101}, \tau_{R-102}, \tau_{R-103}}{Min} \{\phi\}, \qquad \text{(Eq. 1)}$$

$$\phi = [-NPV + (F_{13} - MD_{WPC})\Pr ice_{WPC} + (F_{20} - LM_{WPH})\Pr ice_{WPH} + (F_{31} - LM_{LCPhe})\Pr ice_{LCPhe}]$$

where MD_{WPC}, MD_{WPH} and MD_{LCPhe} are, respectively, market demand constraints for products WPC, WPH and LCPhe; F_{13}, F_{20} and F_{31} are the mass flow rates of the same products; and τ_{R-101}, τ_{R-102} and τ_{R-103} are the operation times of reactors present within biorefinery.

In an attempt to avoid local minima, the three implemented stochastic methods and SQP were applied in parallel for the optimization problem described in Eq. 1. The results were compared in section 3.

2.2. Computation Solution for the Web-based Application

The model architecture of the web application described here employs Java technology (Servlets and Java Server Pages; Sun Microsystems, Inc.). It is basically a Model-View-Controller (MVC) and Layers architecture. MVC has become important in user-interface programming because emphasizes a design that preserves disjoint contents and layout (Deitel and Deitel, 2007). The *Information Layer* provides data to the application (process data for example). Client requests, data processing and contents presentation are in charge of the *Application Logic Layer*. The *Client Layer* is the interface with the user. Applets were used to provide appealing graphical user interface (GUI) through the web. An applet is a program written in Java that is intended not to be run isolated, but rather to be embedded inside another application like a client browser, since it can be in included in an HTML page.

We used Tomcat version 5.5.17 as the application server. Tomcat is an open-source servlet container developed under the Apache license. Its primary purpose is to serve as a reference implementation of the Servlets and Java Server Pages specifications.

The algorithms written in Matlab® 6.5, MathWorks, Inc., are called from the server with the open-source class JMatLink (Müller and Waller, 1999). The server database was MySQL, a popular open-source SQL relational database management system developed, distributed, and supported by MySQL AB. Complete the computational solution for our case study the software Labview® (version 7.1) and the board NI PCI-6052E for data acquisition (National Instruments, Inc.), both installed on a standalone computer. The TCP/IP protocol was applied to data transfer between Labview and application server.

2.3. Hybrid Phenomenological-neural Modeling

The reaction medium within enzymatic reactors present in biorefinery characterizes a complex mixture of amino acids, proteins and peptides. In addition, the variability of its composition is high, depending not only on the upstream process conditions, but also on climate and seasonal factors that affect the original characteristics of the raw material, bovine milk. In this scenario, a mechanistic based model of these reactions would be intractable, leading to high-correlated parameters.

Combination of mass balances with the prediction capability of neural networks may supply hybrid models able to capture the behavior of enzyme kinetics (Pinto et al., 2007; Rivera et al., 2007). Multilayer perceptrons (MLP) neural networks were trained to map the states of each reactor – molar concentrations of components – into the vector of their reaction rates. Differential mass balances of each component were solved numerically, using the MLP predictions of reaction rates. The structure of hybrid approach and the methodology here put forth are described in Pinto et al., 2007.

3. Results and Discussion

Fig. 2a compares performances of the implemented algorithms during optimization of whey biorefinery. Algorithms DE and PSO achieved the best Net Present Values for stationary state simulations while SA only achieved sub-optimal solutions.
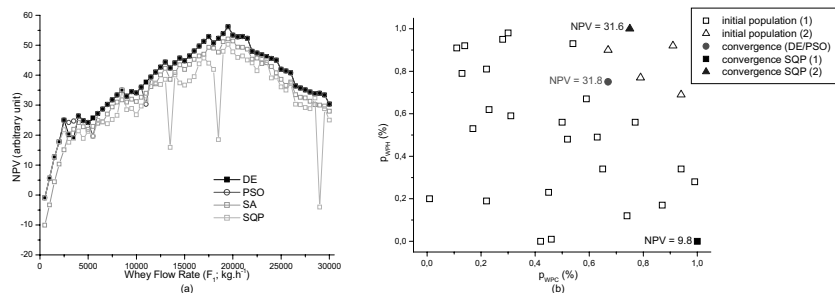
Fig. 2 – Net Present Value as a function of whey flow rate (each whey flow rate defines an optimization problem). **(a)** Performances of algorithms DE, PSO, SA and SQP during the optimization of whey biorefinery (stationary states). **(b)** Performances of DE, PSO and SQP for a whey flow rate of 7500 kg.h^{-1}. In figure, NPV is showed in arbitrary units.

As Simulating Annealing, SQP algorithm also showed poor performances during whey biorefinery optimization. The gradient-based approach only converged to the bound constraints of the variables p_{WPC}, p_{WPH}, τ_{R-101}, τ_{R-102} and τ_{R-103}, while best results achieved by DE and PSO were obtained inside the feasible region. Fig. 2b illustrates this behavior for one case (whey flow rate of 7500 kg.h^{-1}). Initial population of DE and PSO was employed as initial guesses for SQP. Note that while PSO and DE converged to the same result inside the feasible region (the "best" NPV – 31.8), SQP attained two different results in the contours of the same search space (31.6 and 9.8).

Figure 3a presents the web application GUI during the monitoring of a proteolysis to remove phenylalanine from the mixture of peptides, providing LCPhe. An appealing and consistent, real-time updated description of the process was attained. With the hybrid modelling approach, we believe to be possible to follow the course of the enzymatic reactions with real-time accurate predictions, avoiding lengthy off-line HPLC analyses (Pinto et al., 2007).Figure 3b illustrates the control of the pH of the enzymatic hydrolysis of WPC by alcalase®. Possible dynamic time delays caused by the Internet traffic were also overcome by the Internet-based architecture.

The simulation module of the software also allows an economic analysis of the process, including the impact of the process costs (such as enzyme prices and life times, product market demands, etc.). Subject to market demand constraints for the biorefinery products, it is also possible to simulate optimal stationary states of the process using the implement nonlinear search algorithms (DE or PSO for instance).

4. Conclusion

A web-based remote application for economic evaluation, monitoring and control of a whey biorefinery was presented in this paper. In a single integrated environment, is possible to perform real-time operations, such as control and monitoring of operational units, and to simulate stationary states of the biorefinery. The simulation module provides preliminary design of the major equipments, estimative of their fixed and operational costs, economic evaluation of the entire plant and impact of process and market indices variations on process economics. Nonlinear search algorithms (SA, DE and PSO), for process design and simulation were also developed and implemented within the application. With the simulation and optimization modules, the application can act as a decision-support tool for the analyses of different dairy market scenarios.

Hybrid phenomenological-neural models were incorporated as part of the real-time monitoring system, conferring accurate predictions of the components profiles along

reactions. The monitoring of the three enzymatic reactions of the biorefinery, crucial steps of the plant, was an important case study for the web application, which provided an appealing and reliable depiction of the enzymatic reactors.

(a)   (b)

Fig. 3 – Graphical user interfaces (GUI) of the Web application. (a) On-line remote monitoring of the Phe released during hydrolysis with CPA. (b) Remote pH control of WPC production: servo problem.

References

Cabrera-Padilla, RY; Pinto, GA; Giordano, RLC, et. al. 2009, A new conception of enzymatic membrane reactor for the production of whey hydrolysates with low contents of phenylalanine. *Process Biochem*, 44, 4, 269–276.

Clemente, A, 2000, Enzymatic protein hydrolysates in human nutrition. *Trends Food Sci Tech*, 11, 7, 254–262.

Deitel, HM; Deitel, PJ. 2007, *Java How to Program*. 7 ed., Prentice-Hall, New Jersey.

Galvão, CMA; Pinto, GA; Jesus, CDF; et.al. 2009. Producing a phenylalanine-free poll of peptides ater tailored enzymatic hydrolyses of cheese whey. *J Food Eng*, 91, 1, 109–117.

Kennedy, J; Eberhart, RC. 1995, Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks*. Perth, Australia, 4, 1942-1948.

Kirkpatrick, S; Gelatt, CD; Vecchi, MP. 1983, Optimization by simulated annealing. *Science*, 220, 4598, 671–680.

Müller, S; Waller, H. 1999, *Efficient integration of real-time hardware and web based services into Matlab*. Available in http://jmatlink.sourceforge.net/docs/ESS99.pdf.

Pinto, GA; Giordano, RLC; Giordano, RC. 2007, Neural network inference of molar mass distributions of peptides during Tailor-made enzymatic hydrolysis of cheese whey: effects of pH and temperature. *Appl Biochem Biotech*, 143, 2, 142–152.

Pinto, GA; Giordano. RLC; Giordano, RC. 2009, Remote engineering for a cheese whey biorefinery: an Internet-based application for process design, economic analysis, monitoring, and control of multiple plant sites. *Bioproc Biosys Eng*, 32, 1, 67-98.

Rivera, EC; Costa, AC; Andrade, RR; et. al. 2007, Development of adaptive modeling techniques to describe the temperature-dependent kinetics of biotechnological processes. *Biochem Eng J*, 36, 2, 157–166.

Shi, Y; Eberhart, RC. 1998, A modified particle swarm optimizer. In: *Proceedings of the IEEE International Conference on Evolutionary Computation*, Anchorage, Alaska, 69–73.

Siso, MIG. 1996, The biotechnological utilization of cheese whey: a review. *Bioresource Technol*, 57, 1, 1–11.

Storn, R; Price, K. 1997, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J Global Optim*, 11, 4, 341–359.

Optimization of Bioethanol Distillation Process – Evaluation of Different Configurations of the Fermentation Process

Marina O. S. Dias, [a] Tassia L. Junqueira, [a] Rubens Maciel Filho, [a] Maria R. W. Maciel, [a] Carlos E. V. Rossell, [b] Daniel I. P. Atala[c]

[a] *School of Chemical Engineering, State University of Campinas – UNICAMP, P.O. Box 6066, Campinas, SP 13083-970, Brazil*
[b]*Interdisciplinary Center for Energy Planning, State University of Campinas, UNICAMP, P.O. Box 6192, 13400-970, Campinas – SP, Brazil*
[c] *Sugar Cane Technology Center, Fazenda Santo Antônio, 13400-970, Piracicaba-SP, Brazil*

Abstract

Process simulation was used to analyze bioethanol distillation process, which requires a large amount of thermal energy. As it is shown in this study, in the ethanol production process the fermentation stage has a significant impact on energy consumption in the purification step. Thus, alternative configurations in the fermentation and distillation processes were proposed and evaluated. The results showed that vacuum extractive fermentation coupled with triple effect distillation presented the lowest energy demand among the studied configurations.

Keywords: bioethanol, fermentation, distillation, simulation, multiple effect distillation.

1. Introduction

Increase in oil prices and growing international concern about climate change and greenhouse gases emissions have motivated the use of renewable fuels, such as bioethanol, as either a gasoline additive or its complete substitute (Balat et al., 2008). Corn and sugarcane are the feedstock used in the US and in Brazil, respectively, which are the largest ethanol producers in the world. Sugarcane is so far the most efficient raw material for bioethanol production: the consumption of fossil energy during sugarcane processing is much smaller than that of corn (Macedo et al., 2008). Furthermore, optimization of bioethanol production process from sugarcane is still possible, and significant reduction of energy consumption can be achieved through the optimization of fermentation and distillation processes.

Two configurations of the fermentation process are most frequently employed in Brazil: feed-batch and continuous, both with cells recycle. Average ethanol content of the wine in several production units in São Paulo state reaches only 9 % (v/v), varying between 7 and 10 % (v/v). This relatively low ethanol content is a consequence of the inhibitory effects of both ethanol and substrate on the yeast cells, among other factors; as a result, a large consumption of energy is required during the purification step, on which a series of distillation and rectification columns are used.

Several techniques to remove ethanol from fermentation broth have been studied, for instance, Bui et al. proposed selective adsorption on silicalite, active carbon and ZSM-5 and concluded that silicate appeared to be the most suitable adsorbent for this process.

Other alternative is the vacuum extractive fermentation, which consists in the continuous removal of ethanol from the fermentation medium by integrating the fermentation reactor to a vacuum flash evaporator (Silva, 1999 and Atala, 2004). This configuration allows the use of higher substrate concentrations in the reactor, compared to conventional fermentors, and the removal of ethanol from the fermentation media keeps ethanol concentration in the reactor in low levels (around 5 % v/v), which has a low inhibitory effect. In addition, wine removed from the flash evaporator contains a higher ethanol content, what decreases energy consumption during the purification step. In this work, simulations of both conventional and vacuum extractive fermentation processes, as well as the purification step for each case, were carried out using software Aspen Plus. Different distillation columns configurations were studied: conventional, double effect and triple effect. The alternative configurations were evaluated and compared to the conventional ones in terms of energy consumption.

2. Fermentation

2.1. Conventional fermentation

In the conventional fermentation process sugarcane juice is used as raw material, with a concentration of sugars around 200g/L. Sugars are converted into ethanol, CO_2, acetic acid and glycerol. Wine concentration is relatively low (around 8 wt% ethanol).

2.2. Vacuum extractive fermentation

The vacuum extractive fermentation process consists of a fermentation reactor coupled to a vacuum flash evaporator, which allows ethanol produced to be simultaneously removed from the reactor. Thus, ethanol concentration in the reactor remains at low levels (around 8 wt% ethanol), while the wine obtained in the flash chamber contains a concentration equal to approximately 36 wt% ethanol. Because ethanol is continuously removed from the fermentation media, sugarcane juice may have a larger concentration of sugars (around 450 g/L), which is more than two times greater than that of the conventional process. Configuration of this process is depicted in Figure 1.
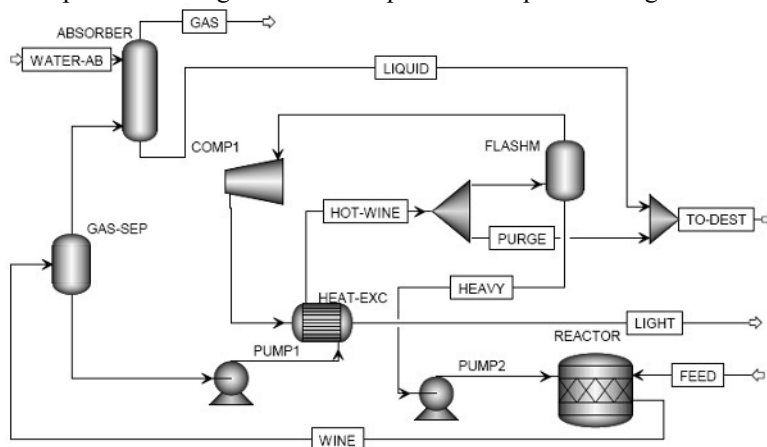

Figure 1 .Configuration of the vacuum extractive fermentation process.

2.3. Simulation data

In both cases, sucrose is almost entirely converted into glucose and fructose (conversion of 99 %); 99.5 % of the reduced sugars are converted into ethanol and CO_2, and the remaining fraction is used as source for the production of by-products (acetic acid and

glycerol), represented by equation 1 (Franceschin et al., 2008):

$$2C_6H_{12}O_6 + H_2O \rightarrow C_2H_4O_2 + 2C_3H_8O_3 + C_2H_5OH + 2CO_2 \qquad (1)$$
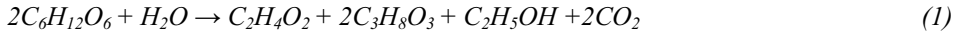
3. Distillation

In the conventional fermentation process, only one ethanol-rich stream (wine) is produced, as opposed to the alternative process, on which two streams are obtained ("light" and "to-dest", comprised by flash vapour phase and a mixture of a reactor purge and alcoholic solution from the absorber, respectively, as shown in Figure 1). Thus, the distillation columns for each case may have different configurations, due to the characteristics of the feed streams.

3.1. Conventional distillation process

The conventional configuration considered in this work is based on a typical industrial scale process, on which a series of distillation (A, A1 and D, located above one another) and rectification columns (B and B1) are employed to produce hydrous bioethanol (around 93 wt% ethanol). In this process, the feed streams are fed to the distillation columns, on which vinasse, vapour and/or liquid phase phlegms and gases are produced: liquid phlegm is a bottom product of column D and vapour phlegm is produced near the top of column A. Phlegms are fed to the rectification columns, producing hydrous ethanol and phlegmasse. All columns work under pressures around atmospheric. This simulation was taken as base for the alternative configurations.

3.2. Double effect distillation process

The double effect process is similar to the conventional configuration, but the distillation columns operate under vacuum (19 – 25 kPa), while rectification columns operate under atmospheric pressure (101 – 135 kPa). In this way, different temperature levels are observed between column A reboiler and column B condenser (65 and 78 ºC, respectively), allowing thermal integration of these equipments and consequently reducing energy consumption on the distillation stage. Configuration of this process applied to the alternative fermentation is depicted in Figure 2.
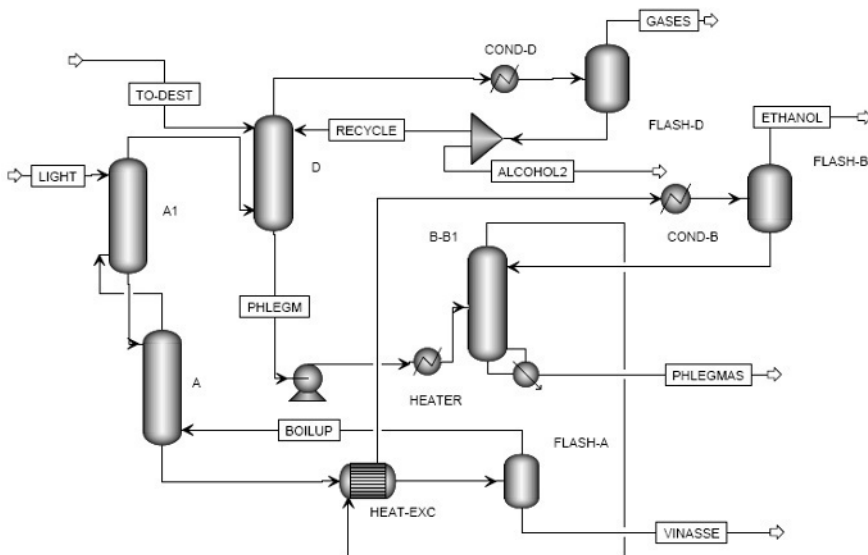

Figure 2.Configuration of the double effect distillation process.

3.3. Triple effect distillation process

In the triple effect configuration, the distillation columns operate under vacuum (19 – 25 kPa), and the liquid phlegm stream produced on column D is split in two: one of them is fed to a rectification column operating under nearly atmospheric pressure (column "B", 70 – 80 kPa) and the other is fed to a rectification column which operates under relatively high pressure (column "B-P", 240 – 250 kPa). Thermal integration between columns reboilers and condensers are possible, as represented on Figure 3.
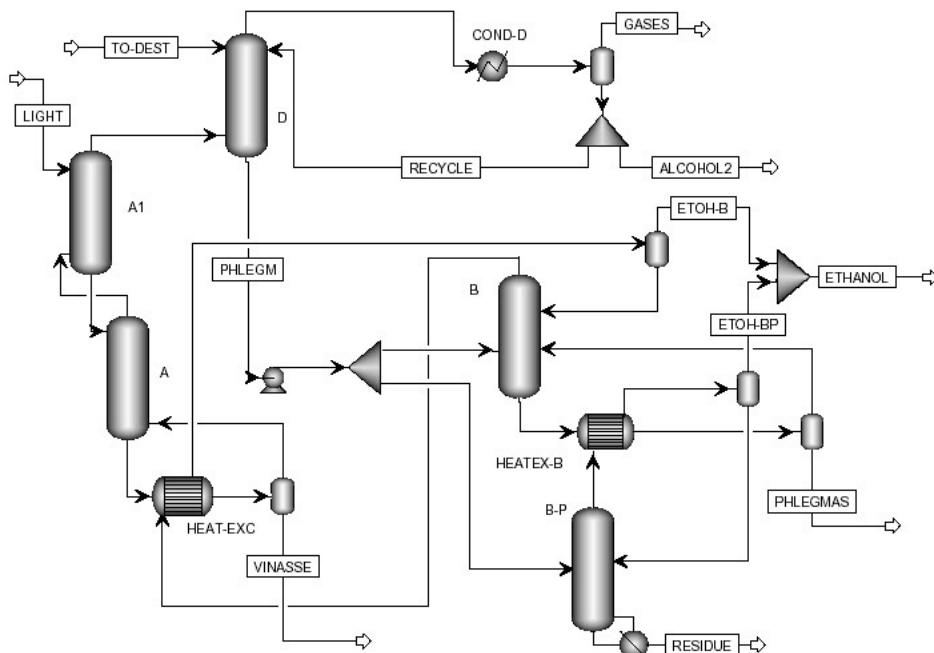

Figure 3. Configuration of the triple effect distillation.

4. Simulations and results

Six different configurations were simulated and evaluated: conventional fermentation and distillation with liquid phlegm only (CFCD), conventional fermentation and distillation with both liquid and vapour phase phlegms (CFCDP), vacuum extractive fermentation and conventional distillation with liquid phlegm only (VFCD), vacuum extractive fermentation and conventional distillation with both liquid and vapour phase phlegms (VFCDP), vacuum extractive fermentation and double effect distillation (VFDD) and vacuum extractive fermentation and triple effect distillation (VFTD). Sobočan and Glavič (2000) also performed simulations of distillation columns for ethanol production and verified that thermally integrated columns provide a reduction in annual costs. In this work, energy consumption of the operations involving heat exchange was evaluated, considering hydrous ethanol production on each case. Results are shown in Table 1, in which heat duty of thermally integrated streams, such as that of the equipments "HEAT-EXC" shown in Figures 2 and 3, is equal to 0. It can be inferred from Table 1 that changes in the distillation process configuration are needed when the alternative fermentation process is considered: the withdrawn of a vapour phase phlegm

on the conventional distillation column does not provide a significant reduction on energy consumption (4 %), as opposed to the conventional fermentation process (18 %).

Table 1. Energy consumption for each process.

| Operation | Energy consumption (kJ/kg hydrous bioethanol) | | | | | |
	CFCD	CFCDP	VFCD	VFCDP	VFDD	VFTD
Reactor cooling	1132	1121	0	0	0	0
Wine heating	3014	2985	0	0	0	0
Column A reboiler	6704	6197	1475	1156	0	0
Column D condenser	6016	4456	2651	2064	2125	2118
Liquid phlegm heating	0	0	19	39	498	0
Column B reboiler	4790	2784	3339	3438	3488	2714
Column B condenser	3669	2720	2346	2727	1304	0
Heating – total	14507	11966	4833	4633	3986	2714
Cooling - total	10817	8297	4997	4791	3429	2118

Since the alternative fermentation process operates under vacuum pressures, the "light" stream must be compressed before being fed to the distillation columns. In the conventional distillation configuration, at least three compressors must be used to raise stream pressure; in the double and triple effect, only one compressor may be used, since the pressure difference between the "light" stream and its feed stage is relatively small. Thus, electricity consumption in the distillation processes increases when the alternative fermentation process is considered: around 600 kJ/kg for the double and triple effect configurations, and around 1200 kJ/kg for the conventional distillation configurations.

Considering exclusively the consumption of thermal energy, the configuration that presents the lowest energy demand is the triple effect configuration (VFTD), providing a reduction in energy consumption equal to 44 % when compared to the VFCD process and 77 %, when compared to the CFCDP process, which is the configuration most commonly employed in Brazilian biorefineries. Nevertheless, the VFTD configuration presents the highest ethanol losses, along with the VFDD process, as shown in Table 2.

Table 2. Ethanol losses on each process.

Operation	CFCD	CFCDP	VFCD	VFCDP	VFDD	VFTD
Ethanol losses (kg/h)	3.6	3.3	1.8	1.7	5.1	5.1
Ethanol losses (%)	10.0	9.2	5.0	4.7	14.2	14.2

In order to reduce ethanol losses in the double and triple effect configuration (VFDD and VFTD), an absorber column may be employed to recover ethanol present in the "GASES" stream, since it accounts for more than 90 % of the losses.

Different stream flows are observed for each process configuration. Because of the increased wine concentration in the vacuum extractive fermentation processes, vinasse volumes are larger in the conventional fermentation process, and phlegmasse (stream "PHLEGMAS" in Figures 2 and 3) is also a significant residue generated in the distillation process. Values of flow for the main process streams are shown in Table 3.

Table 3. Stream flows for each process.

Stream	Stream flows (kg/h)					
	CFCD	CFCDP	VFCD	VFCDP	VFDD	VFTD
FEED	460.7	460.7	169.0	169.0	169.0	169.0
LIGHT	-	-	84.9	84.9	84.9	84.9
TO-DEST	460.7	460.7	69.6	69.6	69.6	69.6
GASES	37.1	37.1	12.4	12.4	17.0	17.0
ALCOHOL2	1.2	0.9	0.6	0.4	0.3	0.3
VINASSE	312.7	319.7	78.1	83.2	22.7	22.7
PHLEGMAS	74.8	67.8	26.5	21.3	81.1	81.0
ETHANOL	35.0	35.3	37.0	37.2	33.4	33.5

5. Conclusions

The combined simulation study of the fermentation and distillation processes for bioethanol production provides important resources that allow the comparison of different process configuration and their impacts on bioethanol losses, energy consumption and residue generation. In this work six process configurations were evaluated, based on conventional and vacuum extractive fermentative processes and conventional, double and triple effect distillation systems for product purification. The optimum configuration regarding energy consumption (the triple effect configuration, VFTD) requires 77% less thermal energy, when compared to conventional fermentation and distillation processes. However, this configuration presents the largest ethanol losses, which may be minimized by the addition of an absorber column, since one of the streams accounts for more than 90 % of the losses.

6. Acknowledgments

References

D.I.P. Atala, 2004, PhD Thesis, School of Food Engineering, State University of Campinas, 2004

M. Balat, H. Balat, C. Öz, 2008, Progress in bioethanol processing, Progress in Energy and Combustion Science, 34,551-573

S. Bui, X. Verykios, R. Mutharasan, 1985, In situ removal of ethanol from fermentation broths.1. Selective adsorption characteristics, Ind. Eng. Chem. Process Des. Dev, 24, 4, 1209-1213

G. Franceschin, A. Zamboni, F. Bezzo, A. Bertucco, 2008, Ethanol from corn: a technical and economical assessment based on different scenarios, Chemical engineering research and design, 86, 488-498

I. C. Macedo, J.E.A. Seabra and J. E. A. R. Silva, 2008, Green house gases emissions in the production and use of ethanol from sugarcane in Brazil: The 2005/2006 averages and a prediction for 2020, Biomass and Bioenergy, 32, 7, 582-595

F.L.H. Silva, M.I. Rodrigues, F. Maugeri, 1999, Dynamic modelling, simulation and optimization of an extractive continuous alcoholic fermentation process, J Chem Tech Biotech, 74,176-182.

G. Sobočan, P. Glavič, 2000, Optimization of ethanol fermentation process design, Applied Thermal Engineering, 20, 529-543

Operational Conditions in Oil Recovery with Blanket Heating in Shallow Reservoirs

Elthon J. R. Medeiros[a], Janusa S. Araújo[a], Tommy A. Pinto[a], Jennys L. M. Barillas[a], Tarcilio V. Dutra Jr[a,b], Wilson da Mata[a,c]

[a]*Programa de Pós-Graduação em Ciência e Engenharia de Petróleo, Universidade Federal do Rio Grande do Norte,Campus Universitário, Natal - RN 59078-970, Brasil*
[b]*Departamento de Engenharia Química, Universidade Federal do Rio Grande do Norte,Campus Universitário, Natal - RN 59078-970, Brasil*
[c]*Departamento de Engenharia Elétrica, Universidade Federal do Rio Grande do Norte,Campus Universitário, Natal - RN 59078-970, Brasil*

ABSTRACT

The most innovative proposals in the field of petroleum research refer to the oil recovery previously considered economically unviable in the market. Included in this premise is low °API oil, also known as heavy oil. Statistics show that, currently, the world reserves could be greatly improved by heavy oils contained in formations with depths between 50 m and 300 m, which are classified as shallow or ultra-shallow reservoirs. Thermal methods have been the most effective alternative for heavy oil recovery, and among them, there is the steam flood, the technique most often applied to reduce oil viscosity. A recently developed method, called Blanket Heating, combines some of the fundamental characteristics of thermal methods, adapting them to the particularities of the shallow reservoirs. This process works introducing steam in horizontal metal conduits, meaning that the heated fluid does not come into direct contact with the formation, working as a classical heat exchanger. The heating occurs indirectly, avoiding problems such as the recovery of large amounts of water and the insertion of excessive volume of steam, especially in cases where the depth is minimal and the overlying pressure is insufficient to contain the fluid. The primary focus of the article is to examine the influence of parameters involved in the operation of blanket heating, in order to find an optimal operational configuration. The choice of the horizontal direction for conduits reflects the need to maintain the greatest possible contact area between conduits and oil reserves. The results show that the blanket heating may be a viable process for heavy oil recovery in extremely shallow formations. They also show that oil recovery can be maximized in proportion to the increase of temperature in the conduits and the number of conduits. The cumulative oil production is reduced when the distance between the conduits and producer wells or between the own conduits is greater. In addition, the results were better when the completion interval of producer wells and the position of the conduits in relation to the vertical are arranged between the center and base of the reservoir. The inner diameter of the conduits was the only parameter that had minimal influence, showing no significant alterations in the production of oil. The study also showed that the blanket heating does not produce significant emissions of steam to the surface, confirming the reduction in the amount of water produced. In the face of the growing importance of heavy oil on the world market, together with the successful exploration of this resource, and knowing the current

applicability of the steam flood as the main alternative to the economic recovery of this type of oil, justify the need to a study that will expand the options to use this method in reservoirs that retain great amount of viscous oil at lower depths.

Keywords: Steam injection, heavy oil, shallow reservoirs, Blanket Heating, computer simulations.

1. INTRODUCTION

It is unlikely to apply the steamflood process successfully in shallow reservoirs, especially the ultra-shallow ones, since the overburden sediment does not provide sufficient barriers against the steam to ensure that the steam does not migrate to the surface, and it conditions high production of sand and the injected fluids. The steam exhaust would reduce the pressure needed to the reservoir, and also could take with itself the lighter fractions of hydrocarbons, causing environmental impacts. A possible solution to these problems is to use methods of indirect heating.

2. BLANKET HEATING

2.1. Method Description

In the Blanket heating process, the steam produced in a conventional generator flows through horizontal conduits to provide heat throughout the formation. The steam contained in conduits avoids the inconvenience of inserting large amounts of fluid directly into reservoir, and help to maintain the temperature through a injected steam control - and the pressure on the conduits. Producer wells are drilled in reservoir to establish the flow of fluids and can also be used to maintain pressure (Osterloh and Jones, 2001).

The conduits must return to the surface so that steam can be reheated in boilers. An alternative would be the steam coming back internally through the conduits. In this case, an external compartment, call annulare, is used to lead the high-quality steam until tubing end, transmitting heat by conduction to the formation, and an ICCT (Insulated Concentric Coiled Tubing), located internally to the annulare, is used so that the fluid return to the boilers.

2.2. Mechanisms Process

The process involves conducting heat - from steam to oil through conduits. In this sense, an alternative material for the conduits would be one able to improve heat conduction, and resistant enough to withstand the pressure and other factors. Once forwarded to the oil, the heat spreads through hydrocarbon by convection and also by conduction (to a lesser extent). Oil in contact with conduits suffer reduction in density and tends to rise, mobilizing the oil located above, inducing vertical flow and easing the thermal energy transfer and also the wells production. The conduits are placed below the oil, to condition the convection process. The process is also privileged by high permeability (vertical and horizontal) of reservoir shallows.

Besides heat transfer, the success of the process is due to the existence of additional process of vaporization or in situ condensation. Steam formed around conduits rises up passing through sediments, and accelerating the heat transfer to lower temperature zones. When steam is finally condensed, liquid will be drained down, filling the warm zone around conduits, where it will be vaporized again (Osterloh and Jones, 2001).

3. MODELING DESCRIPTION

The selected system uses Cartesian coordinates in "i", "j" and "k", directions with dimensions of 60 m x 60 m x 32 m, totaling 4,440 blocks (Fig. 1 A). The conduits were laid out along the axis "j", and correspond to the horizontal columns of blocks with radial refinement, and diameter equal to 5 cm. Only blocks representing conduits were heated at different temperatures in order to evaluate the capability of this configuration to transfer heat to the reservoir in different situations. Fig. 1 B is a top view of the radial refinement applied to conduits blocks.
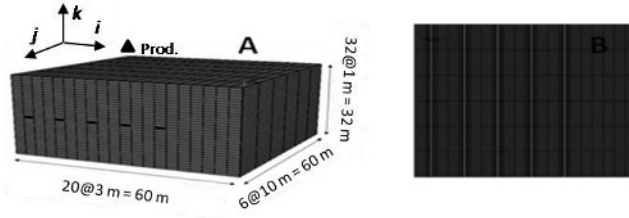

Figure 1.A. 3D representation of base case. B. Conduits (flat "i" -"j").

3.1. Rock-Reservoir Parameters

Table 1 shows reservoir properties, which were based on real reservoirs data and some literature data.

Table 1. Reservoir Properties; Rock Properties.

Reservoir		Rock	
Gas cap - (m)	2	Effective Compressibility of Rock (1/Pa)	$4,4*10^{-7}$
length - i (m)	60	Thermal conductivity of rock (J/m.s.°C)	1,73
Width - j (m)	60	Thermal conductivity of water (J/m.s.°C)	0,61
Thickness - k (m)	32	Thermal conductivity of oil (J/m.s.°C)	0,13
Reservoir depth (m)	50	Thermal conductivity of gas (J/m.s.°C)	0,04
Initial water saturation (%)	30	Horizontal permeability - Kh (mD)	1000
Initial oil saturation (%)	70	Vertical permeability - Kv (mD)	100
Initial reservoir temperature (°C)	37,8	Porosity - φ (%)	28
Original oil in place (m³std)	21509		

Operational parameters used in studied model are shown in Tab. 2. Legend used was: "B", "C" and "T" represents completion and location of conduits, which represent, "Base", "Center" and "Top", respectively. First letter located to the left refer to completion, while the letter to the right is the position of the conduits in relation to the direction "k" (vertical).

Table 2. Levels of Operating Parameters.

Parameter	Levels									
Completion Interval of the well producer and the location of vertical conduits	BC	CB	TT	TB	TC	CC	BB	BT	CT	All B
Distance between conduits (m)	6	9	12	-	-	-	-	-	-	-
Distance between prod. and conduit (m)	3	6	9	-	-	-	-	-	-	-
Quantity of Conduits	1	2	3	4	5	6	7	-	-	-
Radius of the Conduits (cm)	2,5	3,75	5	-	-	-	-	-	-	-
Temperature in the Conduits (°F)	150	200	250	300	350	400	450	500	550	600

4. RESULTS AND DISCUSSION

4.1. Analysis of Conduits Radius

In this section, it was performed an analysis of conduits radius influence on cumulative oil. All parameters shoed in Tab. 2 were considered constants, except the conduits radius, which was studied at three levels (2.5, 3.75 and 5 cm). Fig. 4 shows cumulative oil, cumulative water and cumulative gas for all three radius studied.
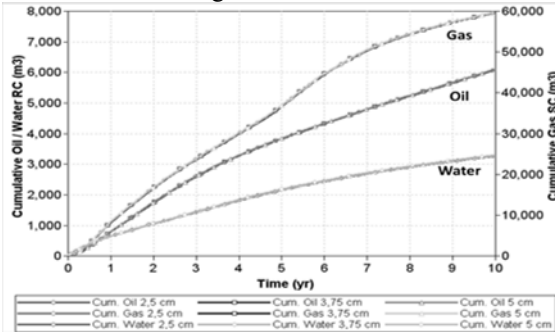

Figure 2. Cumulative Oil Production *versus* Time. Comparative of Radius of the Conduits.

The same behavior can be observed for the three levels as the internal radius does not significantly alter the area of contact of the conduits with the heated oil, changing minimally the heat transfer. Tab. 3 represents a summary of the models with different conduit radius. Oil recovery is given by relation between cumulative oil (CO) and original oil in place volume (OOIP), so OR = CO / OOIP. It can be observed that there was no considerable change in the oil recovery as discussed in the graphics of cumulative production.

Table 3. Oil Recovery e Cumulative Oil at Ended of Production for all Three Conduits Radius.

Radio	CO (m³)	OR (%)
2,50 cm	6066,51	28,20
3,75 cm	6077,56	28,26
5,00 cm	6090,19	28,31

4.2. Analysis of Conduits Temperature

Fig. 3 shows oil recovery (OR) *versus* temperature, for annual production times. Oil recovery increase with time and temperature. A growth of the recovery factor with the rising of the temperature occurs until the fifth year of project, and after that time there is almost no change in the recovery factor at temperatures above 450 °F (232.22 °C). In the eighth year, it was observed that the highest recovery factor occurs to the temperatures between 400 and 500 °F (204.44 and 260 °C).
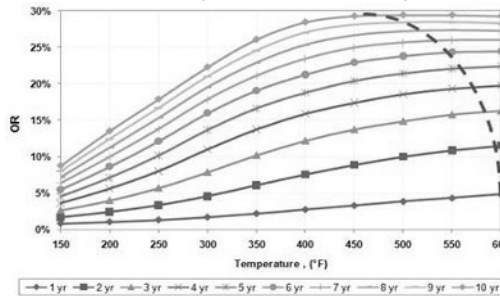

Figure 3. Graph of the Oil Recovery at Each Temperature According to Time.

The red dashed line represents approximately the temperature at which the higher recovery is obtained (qualitative analysis).

4.3. Analysis of Conduits Numbers

It was held in this section a variation in the amount of heating conduits. The analysis was developed using one to six conduits, with fixed distances between them and between the producer well, to check the influence of tubes number on oil recovery. In Tab. 4 there is a growing pattern every years, that is, with more heated conduits, greater cumulated oil production is and it there was no change during the time it was studied.

Table 4. Comparison of Conduits Numbers.

Quantity of Conduits	CO (m3)	OR (%)	Quantity of Conduits	CP (m3)	OR (%)
1	3763	17,50	4	5752	26,74
2	4577	21,28	5	6302	29,30
3	5209	24,22	6	6878	31,98

4.4. Analysis of Distance Between Conduits

In Fig. 4 may be noted that great distances between the heated tubes promote greater cumulative oil due to better distribution of heat inside reservoir. Smaller distances result initially in greater oil rate. In other words, better results are obtained with smaller distances and bigger number of conduits.
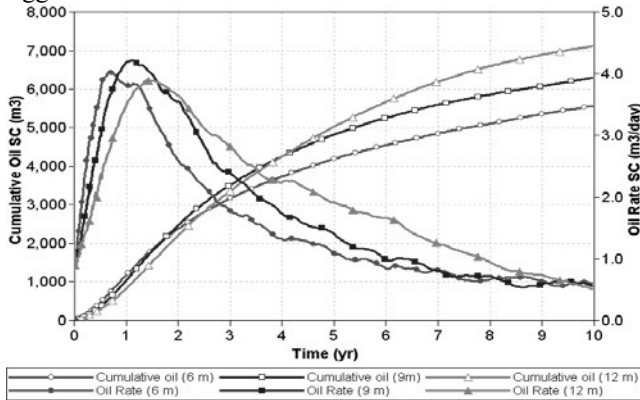

Figure 4. Cumulative Oil and Oil Rat *versus* Time. Distance between Conduits.

4.5. Analysis of Distance between producer and conduit

The closer the conduits are to the producer well the higher is the cumulative oil, resulting on oil viscosity reduction facilitating their mobility, as shown in Tab. 5.

Table 5. Comparison of Distance between Conduits and Producer Well.

Distance between prod. and conduit	CO (m^3)	OR (%)
3 m	6468,86	30,01
6 m	6357,67	29,56
9 m	6066,51	28,20

4.6. Analysis of Completion Interval of the Producer Well and Location of Vertical Conduits

It can be concluded that the best combinations between the completion interval and location of conduits vary over time, between the center and reservoir base. The top had not, at any time considered, a good combination, due to heat loss to the overburden layers of the reservoir, with no significant heating in reservoir bottom (fluids density), and also, less influence of the convective process, because heating occurs on reservoir. The best settings were observed in cases 1, 6 and 7, respectively, producing and heating

at the base (BB), producing at the base and heating the center (BC) and producing and heating in the center (CC) (see Fig 5).
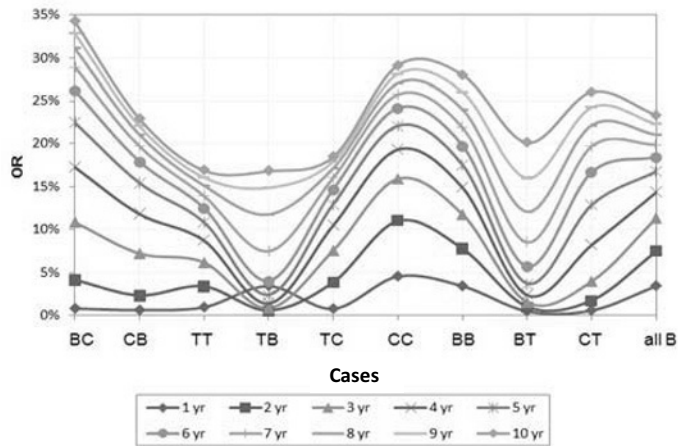

Figure 5. Graph of the Oil Recovery in Each Case in Accordance with the Period.

5. CONCLUSIONS

During the time of study there was increase in the recovery factor with the increasing of temperature in the conduits. The more conduits were heated at high temperatures, the more oil recoveries are obtained.

For the parameters of completion interval and location of conduits, results ruled out the setting at the top due to irrelevant recoveries. The best settings: the base of the reservoir for both parameters (BB), or the base for the completion, with heat in the center (BC).

Oil recovery increases with the reduction of the distance between producer well and conduits, and when the distance between the own conduits is greater, due to better distribution of heated area in the reservoir.

ACKNOWLEDGEMENTS

REFERENCES

1. STARS User's Guide, Advanced process and thermal reservoir simulator, Computer Modelling Group Ltd, 2007.
2. Osterloh, W.T.; Jones, J. Novel thermal process of recovery of extremely shallow heavy oil. SPE 83733, p. 127-134. Março, 2003.
3. Osterloh, W.T.; Jones, J. Process for recovery of extremely shallow heavy oil. In: SPE Western Regional Meeting, SPE 68809, Bakersfield, California, EUA, Março, 2001.
4. Medeiros, E. J. R. Estudo do Aquecimento Geral Indireto como método de recuperação em reservatórios rasos de óleo pesado, Master Thesis, Universidade Federal do Rio Grande do Norte, 226 p., Natal, 2008. (In Portuguese).
5. Dunn-Norman, S.; Gupta, A.; Summers, D. A.; Koederitz, L. F.; Numbere, D. T. Recovery methods for heavy oil in ultra-shallow reservoirs. In: SPE Western Regional/AAPG Pacific Section Joint Meeting, SPE 76710, p. 1-4. Alaska, USA, 20-22, Maio, 2002.

A New Computational Tool for Falling Film Molecular Distillation Performance Prediction

Evandro S. Mallmann[a], Caliane B. B. Costa[a], Maria R. W. Maciel[a] and Rubens Maciel Filho[a]

[a]*School of Chemica Engineering, University of Campinas, UNICAMP, P.O.Box 6066, 13083-970, Campinas – SP, Brazil, engqui.butia@gmail.com, maciel@feq.unicamp.br*

Abstract

In a previous work, a software named DISMOL, was developed in order to simulate molecular distillation. However, due to the restricted access to this tool and the easiness to make use of data bank and physical properties, in the present work, it is proposed the development of general procedure for this highly specific process in the commercial simulator Aspen Plus®. Since no single unit operation is available in the commercial simulator that can appropriately simulate a falling film molecular distillator (MD), the proposal, in a preliminary approach makes use of a sequence of flash vessels in order to accomplish the task of MD simulation. Experimental data of a binary system distillated in a MD are used for the validation of the developed tool. The results indicate the high potential of this tool in representing a MD equipment, making possible the evaluation of different operational policies in conducting this high vacuum distillation to achieve products with the desired quality and properties. The development of the proposed computational tool is an important step for the investigations of operational policies for MD application in oil refining since no industrial data of such equipment is available.

Keywords: Process Model, Molecular Distillation, Aspen, Model Validation

1. Introduction

The petroleum obtained from Brazilian reservoirs is of heavy and ultra-heavy types, presenting high viscosities that make difficult its exploration (oil removal from submarine reservoirs) as well as the refining operations for fuel and other derivatives production. The refining processes currently in activity in the Brazilian refineries do not allow the heavy and ultra-heavy oil processing, so that they are mixed with light oils (of the Arab type) for their processing. In fact this is usually the general strategy adopted worldwide to deal with such type of crude oil. This procedure leads to the dependence of Brazil on oil importation, as well as to the not complete valuation of the oil found in the country. In this context, an alternative to the initial heavy and ultra-heavy oil operation is the use of molecular distillation, a process conducted at high vacuum. Molecular distillation technology, also known as short-path distillation, is an operation for effective separation or purification of products of high molecular weight. It is a special vaporization operation at very low pressures and, consequently, relatively low temperatures, which make this operation also suitable for thermally sensitive products [1], and in particular for the oil refining, avoiding cracking and undesired reactions. Besides that, no external component needs to be introduced into the system in order to perform the operation that can be an advantage for cost reduction in further separation units and in the quality of final product.

Previous experiments [2, 3-4] showed the viability of the MD operation also for heavy and ultra-heavy oil refining. Nevertheless, a deep knowledge of the process is required in order to obtain the desired product streams: small process variations may result in product streams with completely different characteristics [5]. A wide investigation for the development of operational polices that lead to high performance of the molecular distillation (MD) operation for oil refining is necessary. In this context, it is interesting to develop a simulation tool based on commercial simulator since components data bank and facilities to integrate with other unit operation are available.

Bearing all these in mind, the main objective of this work is the development of a simulator of a falling film MD piece of equipment in Aspen Plus® for heavy and ultra-heavy oil processing. The developed tool turns possible an evaluation of the process variables impact on process behavior and characteristics of obtained product streams (residue and distillate). As Aspen Plus ® does not include a MD piece of equipment, it is necessary to choose, among the available unit operations, the one that better suits to the MD parameters. A sequence of flash vessels coupled with process efficiency is selected to represent the process and experimental data are used to validate the model.

2. Methodology, Simulator Development and Validation

In this investigation, data from the binary system Dibutyl phthalate – DBP (278.35 g/mol) and Dibutyl sebacate – DBS (314.14 g/mol) were used for MD process simulation [5]. This binary system was chosen due to its simplicity when compared to the petroleum (multicomponent system), to its high molecular weight and to the availability of published experimental data. Since Aspen Plus® does not possess a MD operation tool, a sequence of flash vessels, operating at equilibrium, were proposed to emulate the molecular distillation. As MD process is ruled by mass transfer rates [6] (non-equilibrium) [7], efficiency considerations must be done in the developed tool in order to fit the simulator outputs to the experimental data. The idea to use a sequence of flashes corrected with process efficiency is a preliminary approach since relatively well established mass-transfer models could be used into Aspen environment to develop the simulator. However, the proposed approach has the advantage to be simpler and be dependent upon only the efficiency values rather than the uncertainties of the mass transfer models. Additionally, to use mass-transfer approach is necessary to identify which could be the more suitable one. This is an important issue to develop the whole simulation methodology and, once it has been considered suitable, the next step should be the use of rate-based approach.

The parameters used for the efficiency and simulator validation were distillation mass ratio (distillate mass/inlet mass) and the molar compositions of distillate and residue, since such data are ready available. The operation temperature was the analyzed parameter, since it is the limiting factor for the great majority of MD applications [8]. These parameters were chosen because independently of the equipment model they are always available and with good accuracy.

2.1. Simulator Development

In the simulations performed, 50 kg/h of an equimolar mixture were fed to the system, operated at 0.001 mmHg. For the MD operation, this process conditions lead to temperature operation of 95.85°C and a distillation mass ratio of 21.2% (i.e., 21.2% of the mass fed to the falling film molecular distillation unit is obtained as distillate), with molar fraction of DBP equal to 0.775 in distillate and 0.429 in residue streams (DISMOL results). The MD simulation first built in Aspen Plus® is shown in Fig. 1, in which, three flash vessels are connected.

The condensed vapor streams obtained with the simulated process depicted in Fig. 1 have, together, a mass rate greater than that obtained with DISMOL simulation. Also, the molar fraction of DBP in "Distill" stream (Fig. 1) is higher than the real MD one (simulated with DISMOL).
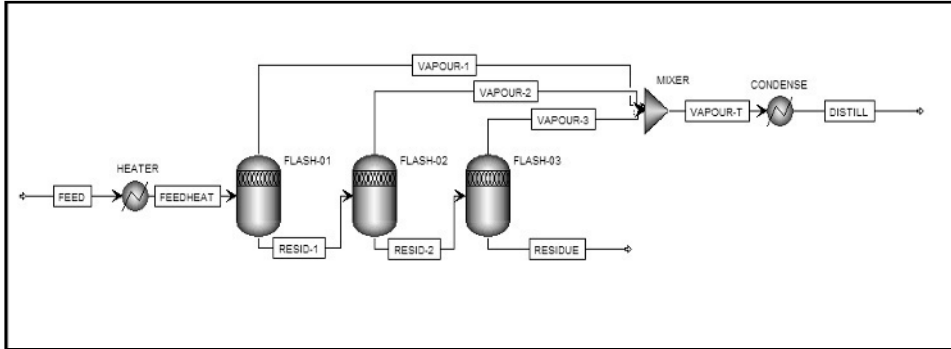

Figure 1: Sequence of flash vessels for DM process simulation in Aspen Plus® Environment.

Furthermore, the operation temperatures of each flash vessel (61.5 °C, 63.0 °C and 64 °C for FLASH-01, FLASH-02 and FLASH-03 respectively) were lower than the real MD temperature for the defined operation pressure. These results are presented in Table 1. The differences in simulation results are due to the fundamental differences with the two unit operations: while a MD process is governed by mass transfer limitations, an ideal flash vessel, as the ones illustrated in Fig. 1, is an equilibrium stage and with an efficiency of 100%.

Table 1. Simulation results for DISMOL and the developed Aspen Plus® tool.

	DISMOL	ASPEN PLUS®	
Distillation mass ratio	0.2120	0.2854	
Distillate DBP molar fraction	0.7750	0.8700	
Residue DBP molar fraction	0.4290	0.3420	
Operation temperature (°C)	95.85	FLASH-01	61.50
		FLASH-02	63.00
		FLASH-03	64.00

The flash sequence leads to a better separation of components since it is an equilibrium operation. However, since MD process is a mass transfer ruled process, it is expected that just one flash vessel may be used (since the more number of flash vessels, the greater the difference between simulation results from non-equilibrium processes results) coupled to an adjusting equation, used to translate flash simulation results to MD process results. Therefore, the final MD tool built in Aspen Plus® possesses just one flash vessel and an adjustment equation to take into consideration that a rate-based process is being emulated by an equilibrium approach.

2.2. Adjusment Tool
Using just one flash vessel, with the selected operation pressure, operation temperature was manipulated in order to evaluate in which operation temperature a distillation mass ratio equals to the one of a real MD process can be obtained. Figs. 2 and 3 show the vapor (which emulates MD distillate) mole fractions and mass flows both for DBP and DBS as a function of temperature. Fig 2 represents the mass flow of the DBP and DBS

components on the vapor stream as a function of flash temperature, while Fig. 3 depicts the molar fraction variation of DBP and DBS on the vapor stream when the flash temperature is changed. From these Figures it is possible to observe that the better separation of system components occurs when flash temperature is 61.5° C. If the flash temperature gets higher, the DBS molar fraction increases and the DBP molar fraction decreases, lowering the vapor purity grade. The condenser turns into liquid phase all the vapor stream coming from the flash vessel. However, in order to obtain the same distillation mass ratio of the DISMOL (21.2 %), the flash equipment has to operate at temperature of 63.1° C. This temperature, nevertheless, results in a higher molar fraction of DBP in the distillate (0.875) and a lower molar fraction of DBP in the residue (0.329).
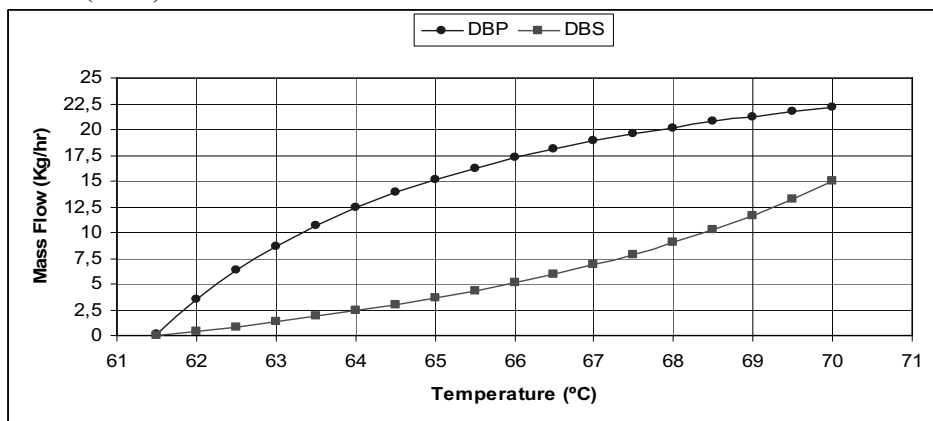
Figure 2. Mass flow of distilled stream as a function of flash vessel operation temperature.
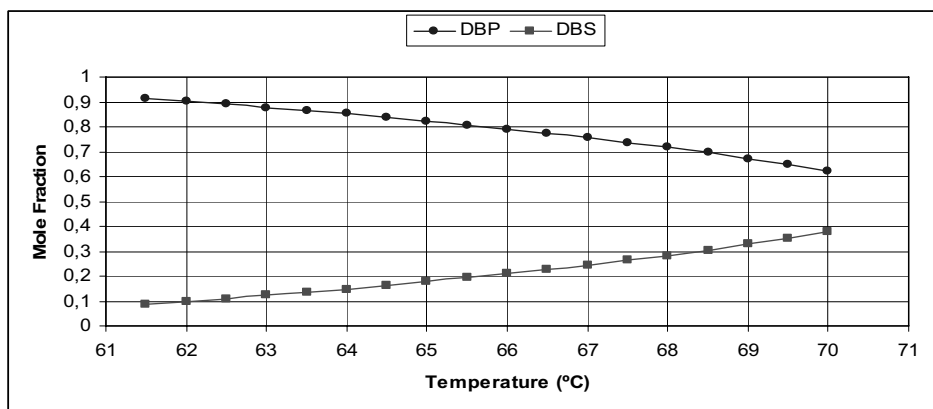
Figure 3. Mole fractions of distilled stream as a function of flash vessel operation temperature.

An adjusting equation, which quantifies an efficiency measurement, is then formulated in order to convert flash results into MD ones, that means to correct the fact that a rate-based process is represent by an equilibrium operation unit. This equation correlates the MD unit operation temperature (given by DISMOL) to the flash operation temperature, so as to provide the same emergent flows composition. The efficiency is calculated, then, dividing flash by DISMOL operation absolute temperatures (Kelvin scale), as depicted in Eq. (1).

$$\eta = \frac{T_{flash}(K)}{T_{DISMOL}(K)} = \frac{336,26}{369} = 0,91 = 91\% \tag{1}$$

The adjusted molar fraction is then calculated multiplying the efficiency factor 'η' to the DBP molar fraction of flash exit vapor stream. The new (corrected) DBP molar fraction in residue is found by mass balance. Table 2 brings DISMOL results compared to the results generated by the developed tool (one flash vessel coupled to an efficiency factor calculated by Eq. (1). In principle, the proposed procedure is general and may be used to any system including multicomponent ones.

Table 2. DISMOL and corrected simulated data (by Eq. 1) for the binary mixture MD.

	DISMOL	ASPEN PLUS®
Distillation mass ratio	0.2120	0.2120
Distillate DBP molar fraction	0.7750	0.7900
Residue DBP molar fraction	0.4290	0.4090

2.3. Heavy Petroleum Simulation

For the evaluation of the heavy oil behavior (multicomponent) in the flash system, the experimental data reported by Santos [9] was used, through the MD. The oil used is called 'gama' petroleum (fantasy name), proceeding from residue at 420ºC+. From the distillate and residue percent data produced in MD, the simulation of the flash system could be carried out, determining different operation temperatures for one system to another, as shown in the Table 3.

Table 3. Comparison between experimental data (MD) and the simulation data (Flash).

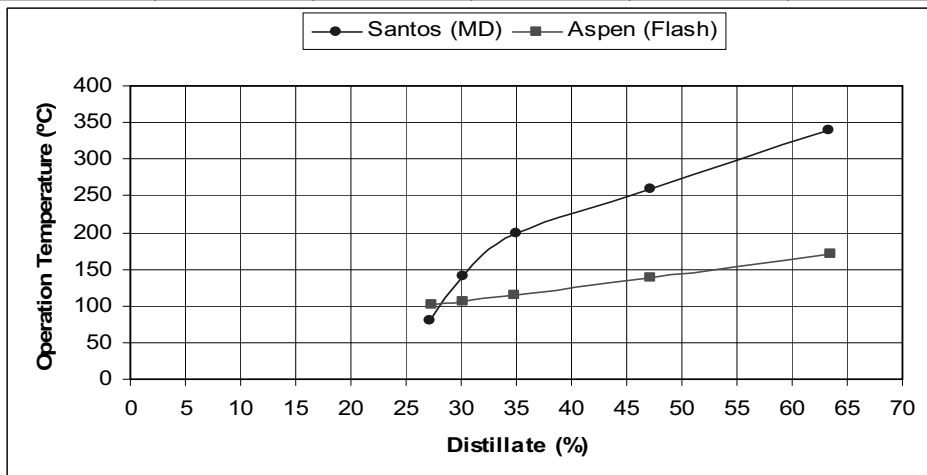| Santos (MD)			Aspen Plus® (Flash)		
Temp.(ºC)	Distill. (%)	Resid. (%)	Temp. (ºC)	Distill. (%)	Resid. (%)
80	27.16	72.84	101	27.30	72.70
140	30.15	69.85	106	30.10	69.90
200	34.94	65.06	115	34.90	65.10
260	47.14	52.86	139	47.22	52.78
340	63.30	36.70	171	63.49	36.51

Figure 4. Behavior of the distillate and the residue percent as function of operation temperatures for MD and flash.

From the experimental values, studies and simulations was made for the behavior evaluation of the flash system operation temperature as function of the simulator response variables (distillate and residue percent), similar to the experimental percents. The difference in the behavior of the curves showed in the Figure 4 is because the very short residence time in the flash system, which needs elevated temperatures to promote the lighter compounds separation. However, the flash needs a variation in the operation temperature lower than MD to produce practically equal distillation indexes.

3. Conclusions

A new tool was developed with the software Aspen Plus® available unit operations to represent MD. As a preliminary approach it is proposed the use of flash operation corrected with process efficiency since MD is a mass transfer limited process. This tool simulates with a good prediction capability literature data for a binary system. The best tool structure was found with the use of one flash vessel coupled to an efficiency factor. The results generated with the developed tool indicated it can be applied to more complex systems, like heavy and ultra-heavy petroleum.

4. Acknowledgements

References

[1] A.T. Erciyes, H. Ishikawa, M. Inuzuka, S. Hiraoka, H. Mori and I. Yamada, I. Chem. E. Symposium Series, 1 (1987) A359

[2] J. A. B. Hernández, L. Z. Linan, A. Jardin, M. R. W. Maciel, R. M. Filho and L. C. M. Oliveira, Rio Oil & Gás Expo and Conference, Rio de Janeiro, Brazil, IBP2712_08 (2008)

[3] E. R. L. Rocha, M. R. W. Maciel, R. M. Filho, C. B. Batistella and L. C. M. Oliveira, Rio Oil & Gás Expo and Conference, Rio de Janeiro, Brazil, IBP2724_08 (2008)

[4] R. S. Rocha, M. R. W. Maciel, R. M. Filho, C. B. Batistella and L. C. M. Oliveira, Rio Oil & Gás Expo and Conference, Rio de Janeiro, Brazil, IBP2730_08 (2008)

[5] C. B. Batistella and M. R. W. Maciel, Computers Chem. Engng., 20 (1996) S19.

[6] J. Lutisan, J. Cvengros, M. Micoslav, Chemical Engeneering Journal, 85 (2002) 225

[7] K.C.D. Hickman, A Review. Chem. Rev., 34 (1943) 51

[8] J. Lutisan, J. Cvengros, M. Micoslav, Chemical Engeneering Journal, 78 (2000) 61

[9] P.S.D. Santos, *Sc. D. These*, Laboratory of Separation Process Development, School of Chemical Engeneering, State University of Campinas, São Paulo, Brasil, 2005

1911

Morphological Population Balance Models for the Dynamic Evolution of Particle Shape and Size Distribution in Protein Crystallization

Jing J. Liu[1,2], Cai Y. Ma[1], Yang D. Hu[2] and Xue Z. Wang[1]*

[1]*Institute of Particle Science and Engineering, University of Leeds, Leeds LS2 9JT, UK*
[2]*School of Chemical Engineering, Ocean University, Qingdao, PR China, 266001*

Abstract

Protein crystallization is known to be affected by many factors and inherently difficult to control. Being able to model the crystal growth, especially at process scale for the population of particles in a reactor rather than for a single particle, will no doubt greatly help the formulation and manufacture of protein crystals. In this paper, a morphological population balance model is presented which has incorporated the crystal shape information into the population balance process model therefore is able to simultaneously simulate the dynamic evolution of shape as well as size for crystals of tetragonal Hen-Egg-White (HEW) lysozyme within a crystallizer. Morphological population balance models require growth kinetics data for each facet, which was obtained from published data in literature for the two identified independent crystallographic faces, {101} and {110}, of HEW lysozyme.

Keywords: morphological population balance model; protein crystallization; HEW lysozyme, crystal shape control

1. Introduction

Protein crystals have shown significant benefits in the delivery of biopharmaceuticals, but the majority of biopharmaceuticals are not yet marketed in crystalline form with only a few exceptions such as insulin. This situation could be partly due to the fact that proteins are difficult to crystallize and the crystallization process is inherently complicated and difficult to control. Proteins, like many other molecules, can be stimulated to form crystals when placed in the appropriate environment. But protein molecules are large, complex and the nutrient solutions for growing protein crystals differ considerably from the simpler nutrient solutions for small molecule crystal growth. The majority of previous work focused on investigations on crystal growth at the molecular and single crystal scales. For the knowledge to be more useful in the design, optimization and control of industrial protein crystallizers, modeling at the process scale to understand the growth behavior of population of the crystals is clearly important.

In modeling crystallization processes, PB provides the dynamic evolution of crystal size distribution in all stages including nucleation and growth, which is also a powerful tool for studying the effects of various operating conditions such as impurity, solvents, cooling rates and supersaturation. As a result, it is considered as a useful tool for product development, process design, optimization and control. Traditionally, crystallization PB modeling has been conducted using a mono-size dimension by defining the size of a particle as the diameter of a sphere having the same volume of the

particle. Such a simplified treatment obviously misses important information about the evolutionary behavior for the morphology of crystals as a population. Motivated by this observation, some researchers have attempted to develop multi-size dimensional PB models for crystallization processes[1,2,3]. However, the only work that can be found in literature on PB modeling of protein crystallization processes [4] did not take into account of the morphology of protein crystals.

Population balance models, in particular multi-dimensional and morphological PB models for protein crystallization processes, could greatly help the understanding of the growth behaviour for the entire population of crystals in the reactor in terms of morphology as well as size distribution. Such knowledge is vital for the successful formulation and manufacture of protein crystals in industrial scale crystallisers. As a first attempt to build morphological population models for protein crystallization processes, we have chosen a protein that has been well studied in literature on the crystal growth behaviour at single crystal scale, the enzyme hen egg white lysozyme.

2. Lysozyme crystallization

Enzyme Hen Egg White (HEW) lysozyme is an enzyme that crystallises in the tetragonal structure with macromolecules in Figure 1. The morphology of HEW lysozyme is dominated by rhomb-octahedron {101} and hexagon-tetrahedron {110} faces (right). Crystal forms are manifested in the external morphology through the multiplicities of tetragonal symmetry with twelve crystal growth surfaces in total, and determined by the relative growth rates of the different symmetry-related faces involved.
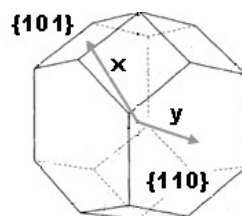

Figure 1. Structure of the tetragonal lysozyme crystal

2.1. Solubility of tetragonal HEW Lysozyme crystals

Solubility is always needed in order to determine the supersaturation at any given operational condition, which is usually used to represent the driving force for crystallisation from solution. Protein solubility can be described as a function of three variables, including pH, precipitant concentrations and temperature. Although efforts were made on measuring solubility, none of the models and techniques is exactly for the operational condition of the current simulation. Therefore in this study we correlated a new solubility equation specifically for the simulated operational condition.

Figure 2. HEW lysozyme solubility

Solubility data plotted for protein HEW lysozyme that was obtained at pH value of 4.5, and different precipitant concentrations of 2.5% from literature, plotted as 'Δ'; 2.5% [5], plotted as 'O'; 2.5% [6], '+'; and 2.0% [7] and 3.0% [8], using blue dashed and dotted curves. Since the simulation condition is pH=4.5 and NaCI concentration of 2.5%, the data of [7] and [8] at 2.0% and 3.0% precipitant concentrations were used to generate data at 2.5% precipitant concentration, plotted as '*'.

It needs to point out that the disparity of the data does not necessarily mean the data is not reliable, because protein crystallisation data is known to have poor repeatability, and minor variation of experiments could give different values. Nevertheless, given the
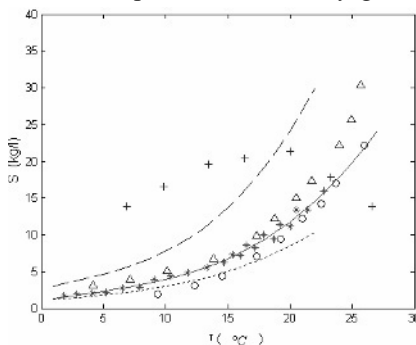
good agreement between the other three sources, one set of the data as shown in Figure 2 as '*' was used to derive the solubility equation with a third-order polynomial:

$$S = 0.0012096 \times T^3 - 0.010496 \times T^2 + 0.26159 \times T + 1.1408 \qquad (1)$$

where S is the solubility of HEW lysozyme crystal, kg/l, and T is temperature, ^oC. The equation is plotted in Figure 2 as a red solid curve.

2.2. Facet growth rate of tetragonal HEW Lysozyme
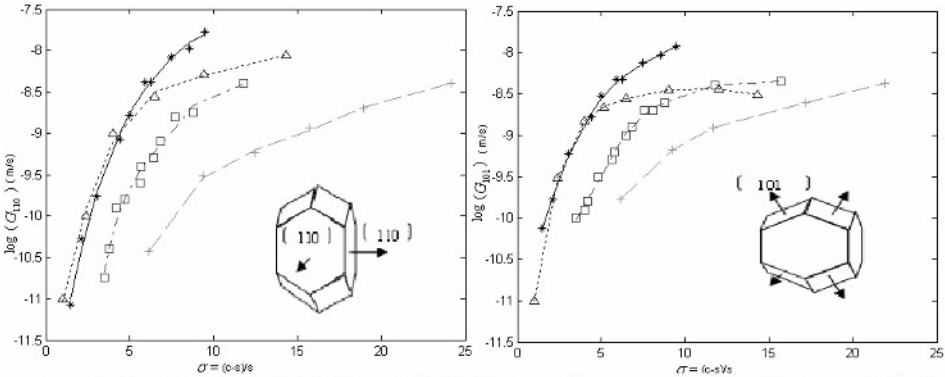
Figure 3. Growth rates of {110} (a), and {101} (b) faces. *– measured at 24 , pH=4.6, and NaCl concentration 3.5%; □– measured at pH=4.5, NaCl concentration 2.5%, and temperature ranging from 12 to 270C; □ – measured at pH=4.5, NaCl concentration 2.5% and temeperature 22 +; obtained under conditions of pH=4.4, NaCl concentration 2.5% and temperature 12.5

There have been experimental studies in literature on growth rates of individual faces of HEW lysozyme crystals. Some of the data is plotted in Figure 3. Although there are some disparities among the data, considering the fact that experiments were conducted in different labs and using different techniques, the growth rate data in Figure 3 is considered comparable and acceptable. The following growth kinetic equations were used to describe the growth rate for each facet:

$$G = k\sigma^n \qquad (2) \qquad\qquad \sigma = (c - s)/s \qquad (3)$$

where G is the facet growth rate m/s, σ relative supersaturation, C the solute concentration *kg/l*, S solubility *kg/l*, k and n are kinetic parameters.

Figure 4. Log – Log plot of the data in Figure 3

Based on the experimental data about facet growth rate, the parameters k and n can be estimated using least square fitting in the log-log scale, as illustrated in Figure 4. Since

n is equal to the slope of the fitting, *k* can be calculated by the intercept of fitting line with Y axis. It was found that the value of exponent *n* calcula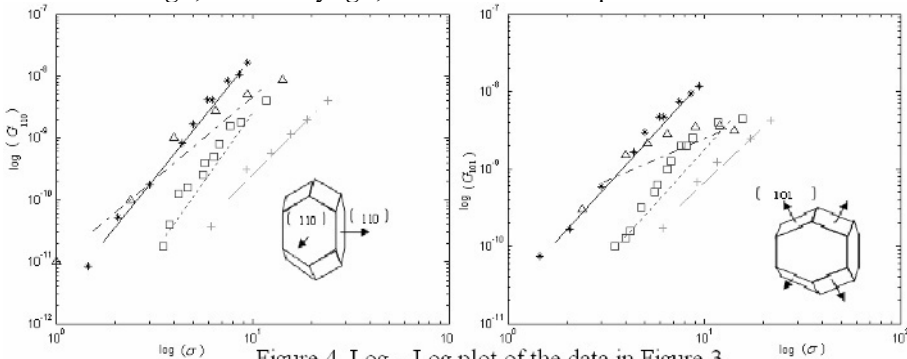ted using the data with dark asterisk of Figure 3 is consistent with the value published [9]. This agreement is proof that the current method for determining the values of the parameters *k* and *n*.

Although values for both *k* and *n* are estimated for all the data plotted, the only data that is similar to our simulation was collected from [10]. Therefore *k* and *n* values obtained from this data were used in the simulation. The *k* and *n* values for the two independent facts are $k_{101} = 5.1 \times 10^{-12}$ m/s, $n_{101} = 2.7$ and $k_{110} = 5.9 \times 10^{-13}$ m/s, $n_{110} = 3.9$.

2.3. Morphology PB equation for Lysozyme crystallization

As stated earlier, due to the symmetry-related faces, the twelve faces of tetragonal lysozyme crystal can be modelled using two independent faces. Therefore the particle shape can be characterized by two internal length coordinates *x*, *y* (Figure 1), which are the distances from the geometric crystal centre to the identified independent faces.

In the current case study, a seeded system with one operating condition (fixed buffer, pH value and salt concentration) is considered and supersaturation is created by reducing the temperature at a constant cooling rate. As is known, in the presence of seed crystals, nucleation can be suppressed not only by different mean seed size and various seed loadings [11], but by keeping the supersaturation within the boundaries where crystal growth occurs, but primary and secondary nucleation do not. So it is appropriate to assume that the amount of solute leaving the solution must be accounted for by crystal growth without nucleated crystals. Under these assumptions, the mathematical formulation for two-dimensional morphological PB modelling can be given by Eq (4), where *f* is crystal size distribution, *x, y* referring to each characteristic length scale, G_{101} and G_{110} refers to the growth rate for the length scale, i.e. the growth rate of face {101} and face {110}. The right-hand side stands for the crystal creation mechanisms, indicating zero new crystals being introduced into this system. This is a simplified treatment, in future research, the net rate of new crystals should be considered.

$$\frac{\partial f(x,y,t)}{\partial t} + \frac{\partial [G_{101}(x,t) f(x,y,t)]}{\partial x} + \frac{\partial [G_{110}(y,t) f(x,y,t)]}{\partial y} = 0 \qquad (4)$$

2.4. Solution algorithm

Solution algorithms for PB equations have been extensively researched in literature. However, it is well-known that the standard first-order schemes give diffusive solutions while the commonly used second-order schemes give spurious oscillations. So it is of great challenge to simulate these equations with extremely sharp distribution of crystal size (1μm to 500 μm), a wide range of length scales and time scale (20 s to 20000 min). High-resolution algorithms have been specifically developed to provide high accuracy while avoiding the numerical diffusion and numerical dispersion associated with other finite difference and finite volume methods.

The high resolution simulation algorithm has proved to be able to provide short computation time and high accuracy in literature. Therefore it is also used in this study. The two-dimensional population balance distributions were obtained via solving the corresponding equations with the number of size classes being 500×400 and the corresponding mesh size being 1×1 μm within a desktop Dell PC (Pentium 4 2.3GHz processor and 3Gb Ram). No spurious oscillations or numerical diffusion occurred, through simulations of HEW lysozyme crystal growth. The numerical stability of the algorithm is assessed using the Courant-Friedrichs-Lewy condition. Since details of the algorithm can be found in literature[12], they will not be repeated here.

3. Results and Discussion

Linear cooling mode is chosen to produce supersaturation in this study and the starting mean normal distances from crystal centre to the face {101} and {110} have been specified as 50 and 15 μm in *x*, *y* directions. It can be seen in Figure 5 that the shape of population distribution is still kept Gaussian-type with the full width at half maximum almost unchanged during protein crystallization process. At the same time, the mean normal distance to the {110} has increased from 50 to 490 μm,

Figure 5. Two-dimensional population distributions

and the mean normal distance to face {101} has changed from 20 to 376 μm.

The growth rates for the two individual faces, G_{110} and G_{101}, have different supersaturation dependences, resulting in change in crystal shape, as illustrated in the inset of Figure 6. when the supersaturation is less than 1.8, the growth rates are quite small and almost no growth in both face directions, which is consistent with the conclusion that cessation of the crystal growth might occur at low supersaturation, like a 'dead zone' where little growth occurs on both faces. For protein crystallization, it is important to control the extent of supersaturation because not only the size, shape and solid-state phase of the product crystals are dependent on the supersaturation profile, but also spuersaturation profile can affect the occurrence of different growth mechanisms. There is a concept named as 'critical supersaturation', which is almost two orders of magnitude higher than that required for a typical inorganic system investigated by experiments. At low supersaturation (high temperature) the distance between opposite {101} pyramids far exceeded the distance between opposite {110} faces. At high supersaturation (low temperature) the reverse occurred. It is worthy noting the crossover of the curves for {110} and {101} faces of HEW lysozyme crystals at $\sigma \approx 6.5$, which is the main reason as crystal habits evolution. The crossover resulted mostly from a decrease in the {101} growth rate at low driving force, which is a consequence of the lattice structure. These phenomena have been explained at a molecular level with Monte Carlo

Figure 6. Two individual faces growth rate for HEW lysozyme

method by Durbin and Feher. Additionally, the dashed line in figure 6 shows the averaged growth rate based on a volume equivalent spherical radius from literature. Some variations can be acceptable due to the implementation of different numerical method for values of kinetic parameters. The inset plot in left up corner showed the proportion of growth rate between face {101} and face {110} with supersaturation. It is not linear and has become less and less with the relative growth rate of two faces,
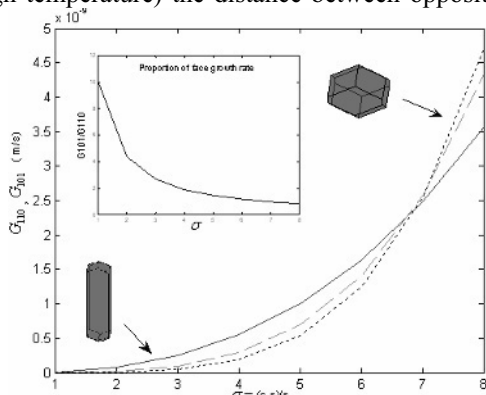
resulting in the changes of crystal habits intrinsically with the shape of face {110} from hexagon to rhombus.

As shown in Figure 7 the crystal shape evolution with time for HEW lysozyme crystals can be observed clearly, from needle-like crystals to plate-like
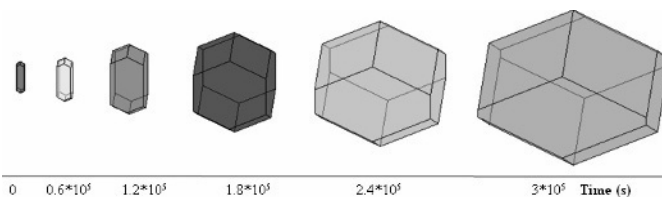

Figure 7. Crystal shape evolution of HEW Lysozyme

crystals during crystallization. Similar results that growth habit changes with σ were obtained by measuring growth rate of {110} and {101} faces of single lysozyme crystal[9]. However, a more quantitative comparison between modelling and these experimental data is not possible for the different operating conditions. The difference of crystal shapes presented here indicates the necessity of using the morphological PB model to capture the habits evolution of protein crystals.

4. Conclusion

Although many different experimental techniques have been used to study protein crystallization, the majority has focused on fundamental investigations at the molecular and single crystal levels. Studies on modelling the protein crystal growth behaviour using population balance equations are still scarce and limited to a size definition for a particle as its volume equivalent diameter of a sphere, neglecting crystal morphology information. The morphological population balance model presented in this work for protein crystallization is able to simulate the dynamic evolution of crystal shape as well as size, providing a potentially useful tool for studying the growth behaviour for the whole population of crystals in a crystalliser. The methodology for protein crystallization process modelling has been validated with a HEW lysozyme protein as an example using literature data.

References

1. Ma, C.Y and X.Z. Wang, AIChE J, 2008, **54**, 2321-2334.
2. Ma, C.Y., X.Z. Wang and K.J. Roberts, AIChE J, 2008, **54**, 209-222.
3. Puel, F., G. Fevotte, and J.P. Klein, Chem Eng Sci, 2003, **58**:3715 - 3729
4. Shi, D., P. Mhaskar, N.H. El-Farra, and P.D. Christofides, Nanotech, 2005, **16**: S562-S574.
5. Rosenberger, F., S.B. Howard, J.W. Sowers, and T.A. Nyce, J Cry Grow, 1993, **129**: 1-12.
6. Howard, S.B., P.J. Twigg, J.K. Baird and E.J. Meehan, J Cry Grow, 1988, **90**: 94-104.
7. Cacioppo, E. and M.L. Pusey. J Crystal Growth, 1991. **114**: 286-292.
8. Forsythe, E.L., R.A. Judge, and M.L. J Chem Eng Data, 1999, **44**: 637-640.
9. Durbin, S.D. and G. Feher. J Crystal Growth, 1986, **76**(3): 583-592.
10. Forsythe, E.F. and M.L. Pusey,. Acta Cryst., 1994, **D50**: 614-619.
11. Hojjati, H. and S. Rohani,. Chem Eng Processing, 2005, **44**: 949-957.
12. Ma, D.L., D.K. Tafti, and R.D. Braatz, InD Eng Chem Res, 2002. **41**: 6217-6223.

Pareto Optimization of an Industrial Ecosystem: Sustainability Maximization

Juliana G. M.-S. Monteiro[a], Patricia A. C. Silva[a], Ofélia Q. F. Araújo[a], José L. de Medeiros[a]

[a]*Dept of Chemical Engineering, Federal University of Rio de Janeiro, Av. Horácio Macedo, n. 2030, Centro de Tecnologia, Bloco E, sala 209, Rio de Janeiro, 21941-909, Brazil*

Abstract

This work aims to design an Industrial Ecosystem for sequestrating of both CO_2 and glycerol in a Chemical Complex with 15 integrated processes to produce: methanol, ethylene oxide, ammonia, urea, dimethyl carbonate, ethylene glycol, glycerol carbonate, β-carotene, 1,2-propanediol and olefins. The Complex is simulated using HYSYS (AspenTech). Processes environmental impact (EI) is calculated using the Waste Reduction Algorithm, while Profit (P) is estimated using classic correlations. The simulation environment is guided by a Matlab (Mathworks) code, which communicates with HYSYS. The performance objective is granting maximum process sustainability, which means finding a compromise between P maximization and EI minimization. Sustainability maximization is understood as a multi-criterial optimization problem, addressed by means of the Pareto optimization methodology for trading off P vs. EI. The general idea behind Pareto optimization is to find a set of solutions in which P can only be improved by compromising on EI and vice-versa. In this work, given the high CPU time demanded for simulating the ecological complex, a procedure for screening the alternatives of CO_2 allocation in the complex is investigated and a reduced Pareto set for the proposed Ecological Chemical Complex is presented.

Keywords: Pareto optimization, sustainability, CO_2 sequestration, process design.

1. Introduction

Global climate change, caused mainly by the increase of CO_2 concentration in atmosphere, is a major concern of the Chemical Industry. Additionally, if biodiesel is to substitute petroleum diesel, glycerol might become a management/environmental problem, since it's generated as a by-product in a 10% (w/w) ratio. The Chemical Industry is currently focusing on environmental performance of processes and products, trying to fit its operation to prevent pollution, using management tools such as Life Cycle Analysis and applying Sustainability Metrics. This work approaches the conceptual design of an Industrial Ecosystem for mitigating the environmental issues attaining chemical and biochemical sequestration of both CO_2 and glycerol to produce: methanol, ethylene oxide (EO), ammonia, urea, dimethyl carbonate (DMC), ethylene glycol (EG), glycerol carbonate (GC), beta-carotene, 1,2-propanediol (1,2-PD), ethylene carbonate (EC) and olefins. Furthermore, a procedure of preliminary screening of the optimal operating region is presented.

2. Chemical Complex Design and Optimization

2.1. Chemical Complex Conceptual Design

2.1.1. Structure

The Chemical Complex proposed in this work is composed by 15 integrated processes, as illustrated in Figure 1. The objective of this complex is to optimize CO_2 chemical and biochemical reuse, by setting the most sustainable CO_2 allocation within the system.

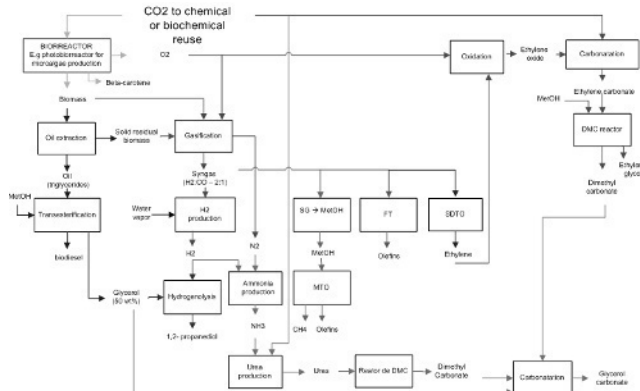

Figure 1: Chemical Complex structure

The processes considered are shortly described as follows: (a) **microalgae biomass and β-carotene production**: CO_2 is biofixed by *Dunaliella salina* cultivation in a photobioreactor. The biomass is then harvested and the β-carotene is separated. The process is simulated using technical data from Chisti (2007) and Araújo *et al*. (2008); (b) **biomass gasification**: biomass is oxidized by O_2 and H_2O to generate syngas with CO:H_2 ratio = 1:2 (Olofsson *et al*., 2005); (c) **biodiesel production**: lipids extracted from biomass react with methanol to produce biodiesel. Glycerol is generated as by-product, as described in González *et al*. (1998); (d) **1,2-PD production**: in a first step, glycerol is dehydrated to acetol; then acetol is hydrogenated to 1,2-PD (Dasaria et al., 2005); (e) **hydrogen production**: vapor reacts with syngas (water-gas shift reaction), generating H_2 and CO_2; (f) **ammonia production**: N_2 from biomass gasification reacts with H_2; (g) **urea production**: ammonia reacts with CO_2 to produce urea (Baal and Lemmen, 2003); (h) **methanol production**: CO and H_2 react under high pressure (7,5 MPa) to form methanol; (i) **methanol-to-olefins (MTO)**: methanol is converted into light olefins, as described by Nouri and Tillman (2005); (j) **syngas-to-DME-to-olefins (SDTO)**: this is a modification of the MTO process, proposed by Cai *et al.* (1995): first, syngas is converted into dimethyl-ether (DME) and then DME is converted to hydrocarbons. The advantage of SDTO route is that syngas conversion to DME is higher than syngas conversion to methanol; (k) **Fisher-Tropsch (FT) synthesis**: FT reaction is modeled according to the Anderson–Schultz–Florey (ASF) model (Trepanier *et al*., 2009). The chain growth probability value was optimized in order to maximize FT selectivity to light olefins; (l) **EO production**: ethylene and oxygen react to produce EO, CO_2 and water. (Coombs et al., 1997); (m) **DMC production from EO**: EO reacts with CO_2, forming EC. EC then reacts with methanol to generate DMC and EG (Cui *et al.*, 2004); (n) **DMC production from urea**: methanol and urea reacts forming methyl carbamate, which reacts with methanol to generate DMC and NH_3 (Wang *et al*., 2007); (o) **GC production**: DMC reacts with glycerol to form GC (Rokicki *et al*., 2005).

The thermodynamic equilibriums of the reactions of processes (b), (e), (f) and are calculated by HYSYS.

2.1.2. Modeling

Casavant and Côtè (2004) ascertain the benefits of using chemical process simulation to design industrial ecosystems, a practice that allows "the design of material exchanges and integrated waste treatment to reduce environmental impact; design facilities that maximize energy efficiency; and design facilities that conserve material use". However, rigorous simulation of a Chemical Complex demands a high computational effort for the convergence of all the recycles involved. In the addressed problem, the simulation flowsheet has 228 streams and 115 unit operations. The optimization of the entire complex demands the simulation to be solved several times, raising the computational effort beyond practicability boundary.

In the adopted procedure, the processes are first rigorously simulated in separate, using HYSYS. Kinetic data (whenever available) are used for modeling the reactions, and reactors (PFRs and CSTRs) and distillations columns are properly sized. These preliminary rigorous simulations allow the calculation of technical indexes such as the conversion of a reactant in a given condition, or the recovery of a product in a distillation column. The Chemical Complex is then simulated using simpler operations models, yet employing the technical indexes calculated and reaction conditions. For instance, PFR reactors are replaced by Conversion reactors, and columns are replaced by HYSYS unit operation block called "component splitter", a decision that accelerates the flowsheet convergence, significantly reducing CPU simulation time. The designed system presents several integrations between material and energy flows: mixers and tees dictate the extent of material integration between processes. The energy integration is calculated considering that the utilities flows are linked into a heat exchangers network, as in Bulasara *et al.* (2008). Additionally, heat exchangers are installed within each process whenever energy integration between process flows is possible.

2.2. Procedure for Chemical Complex Performance Analysis

The performance of the Complex is analyzed by investigating its Sustainability, herein understood as a two-dimensional indicator, considering both economical and environmental aspects. The analysis requires that the Profit (P) and the Environmental Impact (EI) of the Chemical Complex be calculated for the evaluated cases.

The Environmental Impact (EI) Function is calculated using the Waste Reduction Algorithm (WAR), which characterizes sustainability with an index that measures Potential Environmental Impacts (PEI) associated to each process emissions. Technical emission factors are obtained from literature or estimated using HYSYS, in terms of quantity of pollutant released for quantity of product obtained. The greenhouse gas emissions due to utilities demand in operations are calculated as a function of the energy inputs/outputs of the equipment used in each plant. The facilities are considered to employ natural gas as fuel. P of each process is calculated according to Equation 1 (Knapp and Doherty, 1990). ISBL is calculated using Douglas' (1988) correlations; utilities costs are summarized in Table 1; the plants operate for 7200h/y; the Marshal and Swift index (M&S) used is 1399, referring to the year 2007; ISBL for urea production plant is estimated from the values reported by van Baal and Lemmen (2003).

$$P = Revenue - ISBL - (Raw\ Material + Utilities) \qquad (1)$$

Table 1: Cost of Equipments (Douglas, 1988) and Utilities.

Cost of Utilities	
Vapor	USD 6.98/t
Water	USD 6.90^{-3}/m³
Electrical Energy	USD 0.43/kWh

Both P and EI are affected by material and energy integration decisions, often in different ways – since a more profitable process may have more EI and/or be more energy intensive than a less profitable one. Therefore, the mixers and tees of the process are the keys to investigate various situations, allowing the decisions of whether raw material to a given process is obtained within the system (in closed loops), or from outside system's boundaries, or in a combined internal and external solutions. The same applies to process products: is it better to sell a product (send it outside the system) or to use it as reactant within the system?

Normally, as pointed out by Gibbs and Deutz (2007), cycling of materials and energy is understood as a problem of multi-objective minimization of waste production, energy consumption and raw material consumption. Herein, the focus is given to P and EI, an approach that should include cases in which, for instance, a rise in energy consumption due to higher flows in a given process actually leads to both P rising and EI lowering. After the connections between processes are made and process constraints are applied, the Chemical Complex flowsheet remains with high dimensionality: 15 flow ratios (in 11 tees) are to be set in order to solve a flowsheet case, as shown in Table 2.

Table 2: Chemical Complex degree of freedom

TEE op	Inlet	Outlets	Flow ratios
T1	syngas	to H2 production	*to be set (1)*
		to MetOH production	*to be set (2)*
		to FT	*to be set (3)*
		to DME production	calculated by difference
T2	D. salina	to gasification	*to be set (4)*
		to biodiesel production	calculated by difference
T3	Methanol produced	to DMC production (from urea)	adjusted to set MetOH:urea ratio = 2:1 in DMC production
		to MTO	*to be set (5)*
		to be sold	*to be set (6)*
		to DMC production (from EO)	adjusted to set MetOH:EO ratio = 2:1 in DMC production
		Methanol makeup	calculated by difference
T4	CO$_2$ from EO production	recycle	*to be set (7)*
		CO$_2$ purge	calculated by difference
T5	Methanol	to reactor 1 (DMC from urea)	*to be set (8)*
		to reactor 2 (DMC from urea)	calculated by difference
T6	Glicerol produced	to be discharged	*to be set (9)*
		to 1,2-PD production	*to be set (10)*
		to GC production	calculated by difference
T7	Methanol	MetOH purge	*to be set (11)*
		MetOH recycled	calculated by difference
T8	DMC produced	to be sold	*to be set (12)*
		to GC production	calculated by difference
T9	EO produced	to DMC production	calculated by difference
		to be sold	*to be set (13)*
T10	NH$_3$ produced	to urea production	calculated by difference
		to be sold	*to be set (14)*
T11	urea produced	to DMC production	calculated by difference
		to be sold	*to be set (15)*

Flow ratios must respect the following constraints: each value must be contained in the [0,1] interval and the sum of flow ratios of the same tee must equal 1. The number of combinations of flow ratios sets is explosive. Therefore, MATLAB is connected to HYSYS and set to automatically generate random values for each flow ratio, with the given constrains, and save each feasible case. Cases' feasibility depends on HYSYS's capacity of solving the flowsheet for the proposed flow ratio values. The solutions generated by HYSYS are plotted in a P vs. EI graphic, for mapping the Chemical Complex Sustainability performance. The multi-criterial optimization problem of Sustainability maximization is then addressed by means of the Pareto optimization methodology. The general idea behind Pareto optimization is to find a set of solutions in which P can only be improved by compromising on EI and vice-versa.

An initial set of converged flowsheet solutions, 25 cases in the present work, is analyzed for the calculation of the linear dependence of P and EI to each flow ratio. The search for solutions within the Pareto Set is oriented to progressively vary only the most impacting flow ratios, progressively reducing the pool of flow ratios "to be set".

3. Pareto Set Results and Conclusion

Figure 2 shows the map of feasible solutions for the Chemical Complex. Both P and EI are reported in relative terms, divided by the quantity of CO_2 sequestrated in each case. The Pareto set points are market with a red circled and joined by a line – the Pareto frontier. Every point above the Pareto frontier is dominated by a point of the frontier, while the region below the frontier is unfeasible. In a previous work, Monteiro et al. (2009) presented a 2-D Sustainability Function (SF) for process sustainability assessment, as shown in equations 2 and 3:

$$SF = \omega_P P - \omega_{EI} EI_m \qquad (2)$$

$$EI_m = EI \frac{M_{CO_2}}{M_{product}} \qquad (3)$$

where: SF = sustainability function; P = profit; EI_m = modified environmental impact; ω_k = weighting factor associated to function k; EI = environmental impact; M_{CO2} = mass of CO_2-equivalent emitted; $M_{product}$ = mass of product obtained.

This metric was employed then for choosing the best eco-technology between two alternatives. As discussed in the referred work, the ratio ω_P/ω_{EI} is the relevant measure of such metric. This metric is applied to the cases herein studied varying ω_P from 0 to 10 and keeping ω_{EI} constant and equal to 1. The points with maximal SF coincide with the points of the Pareto frontier, meaning that the Sustainability Function can be used as objective function for Sustainability optimization.
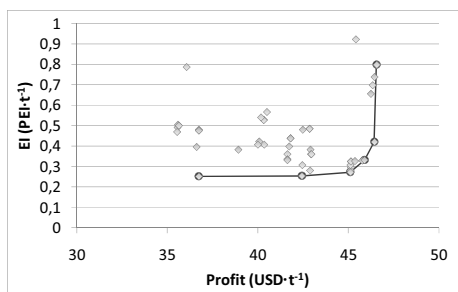

Figure 2: Pareto set (SF maximization solutions)

References

Y. Chisti, 2007, Biodiesel from microalgae, Biotechnology Advances, 25, 294-306.

O. Q. F. Araújo, C. N., Gobbi, R. M. Chaloub, M. A. Z., Coelho, 2008, Assasment of the impact of salinity and irradiance on the combined carbon dioxide sequestration and carotenoids production by Dunaliella salina: a mathematical model. Biotechnology and Bioengineering, Vol. xxx, No. xxx, DOI 10.1002/bit.22079.

I. Olofsson, A. Nordin, U. Söderlind, 2005, Initial review and evaluation of process technologies and systems suitable for cost-efficient medium-scale gasification for biomass to liquid fuels, ETPC Report, University Umea and Mid Swedish University, Sweden.

M. J. I. González, A. R. Medina, E. M. Grima, A. G. Giménez, M. Crastens, L. E. Cerdána, 1999, Optimization of fatty acid extraction from Phaeodactylum tricornutu UTEX 640 biomass, JAOCS, 12.

M.A. Dasaria, P. Kiatsimkula, W.R. Sutterlinb, G.J., Suppes, 2005, Low-pressure hydrogenolysis of glycerol to propylene glycol, Apllied Catalysis, 281, 225-231

H. van Baal, W. Lemmen, 2003, The advantages of a large capacity single line urea plant, Stamicarbon bv Technical Report. Available at http://www.stamicarbon.com/publi_presen/brochures/_en/index.htm

S. Nouri, A.-M. Tillman, 2005, Evaluating synthesis gas based biomass to plastics (BTP) technologies, Chalmers University of Technology, Göteborg.

G. Cai, Z. Liu, R. Shi, C. He, C. Yang, C. Sun, Y. Chang, 1995, Light alkenes from syngas via dimethyl ether, Applied Catalysis, 125, 29-38.

M. A. Trepanier, A. K.Tavasoli, N. Dalai, Abatzoglou, 2009, Co, Ru and K loadings effects on the activity and selectivity of carbon nanotubes supported cobalt catalyst in Fischer–Tropsch synthesis. Applied Catalysis A: General. 353. 193–202.

J. Coombs, K. Daniel, L. Palombo, 1997, Celanese Clearlake Ethylene Oxide Reactor Revamp, JChem Inc. Available at http://www.owlnet.rice.edu/~ceng403/ethox97.html

H. Cui, T. Wang, F.Wang, C. Gu, P. Wang, Y. Dai, 2004, Kinetic Study on the One-Pot Synthesis of Dimethyl Carbonate in Supercritical CO_2 Conditions, Ind. Eng. Chem. Res., v. 43, p. 7732-7739.

G. Rokicki, P. Rakoczy, P. Parzuchowski, M. Sobiecki, 2005, Hyperbranched aliphatic polyethers obtained from environmentally benign monomer: glycerol carbonate; Green Chem., 7, 529–539

T. E. Casavant, R. P. Côté, 2004, Using chemical process simulation to design industrial ecosystems, Journal of Cleaner Production 12, 901–908.

V. K. Bulasara, R. Uppaluri, A. K. Ghoshal, 2008, Revamp study of crude distillation unit heat exchanger network: Energy integration potential of delayed coking unit free hot streams. Applied Thermal Engineering. Article in press.

J.P. Knapp, M.F. Doherty, 1990, Thermal Integration of Homogeneous Azeotropic Distillation Sequences, AIChE J. vol. 36, n.7, 969-984.

J. M. Douglas, 1998, Conceptual Design Of Chemical Processes. McGraw-Hill.

D. Gibbs, P. Deutz, 2007, Reflections on implementing industrial ecology through eco-industrial park development, Journal of Cleaner Production, 15, 1683-1695.

J. G. M-S. Monteiro, O. Q. F. Araújo, J. L. Medeiros, 2009, Sustainability Metrics for Eco-Technologies Assessment, Part II: Life Cycle Analysis. Clean Technologies and Environmental Policy. Article in press.

F. Wang, N. Zhao, J. Li, W. Zhao, F. Xiao, W. Wei, Y. Sun, 2007, Modeling of the Catalytic Distillation Process for the Synthesis of Dimethyl Carbonate by Urea Methanolysis Method. Ind. Eng. Chem. Res, 46, pp. 8972-8979.

1923

Exergy and Sustainable Development for Chemical Industry Revisited

Moises Teles dos Santos,[a] Song Won Park[a]

[a]*LSCP/CESQ - Department of Chemical Engineering, Polytechnic School - University of São Paulo Av. Prof. Luciano Gualberto, n. 380, trav. 3, CEP 05508-900, São Paulo - SP, Brazil. e-mail address: sonwpark@usp.br*

Abstract

It becomes necessary the standardization of environmental sustainability indicators for energy conversion systems as well as the measurement of the impact of their emissions on the environment. The exergy concept has been evaluated as a useful tool for this purpose. It is based on thermodynamic criteria and reveals the transformation capacity of a given energy carrier in the environment and the consumption and depletion flows of the useful resources available on nature. Natural and artificial systems are supported by energy useful potential (exergy) and the evaluation of this exergy consumption in relation to natural resources is an indication of process sustainability. The aim of this work is review the environmental sustainability concepts from the perspective of exergy analysis, show the contribution of exergy to the Life Cycle Analysis and clarify the relations between exergy and environmental impacts measurements. In the final section, we illustrate with two real industrial examples.

Keywords: exergy, energy, sustainability, natural resources.

1. Introduction

The search for sustainable development must be based on three pillars: economic viability, social concerns and ecological issues. Therefore, despite the fact that sustainability is a multidisciplinary concept involving different areas, process engineering plays a fundamental role in this context as it can promote more sustainable process and products (Bakshi and Fiksel, 2003). Application of this concept, however, has claimed for development of metrics: criteria and sustainability indicators definitions that can be used for measurement, evaluation and comparison of industrial activities impacts on the environment (Yi *et al*., 2004; Ness *et al*., 2007). In this context, Green Engineering is the design, discovery and implementation of engineering solutions for sustainability in a multi-scale approach involving molecular, product, process and systems level (Anastas; Zimmerman, 2006; Garcia-Serna *et al*., 2007; Charpentier, 2008). Anastas and Zimmerman (2006) stated the so-called The Twelve Principles of Green Engineering and the goal of this review is highlight how the use of Second Law of Thermodynamics by means of exergy concept can be used to address some of these principles (as listed below), besides its usefulness as a metric for sustainability.

- **Principle 4**: System components should be designed to maximize mass, energy and temporal efficiency.

- **Principle 5**: System components should be output pulled rather than input pushed through the use of energy and materials.
- **Principle 6**: Embedded entropy and complexity must be viewed as an investment when making decision choices on recycle, reuse or beneficial disposition.
- **Principle 10**: Design of processes and systems must include integration of interconnectivity with available energy and material flows.

2. Exergy

2.1. Introduction

Energy is always a conserved quantity. However, it becomes less available to perform useful work and drive processes. This useful work is quantifiable by means of exergy. The exergy is not conserved and this fact can be used for evaluate processes efficiency and consumption of natural resources. Exergy provides qualitative measure of material and energy streams and can identify process irreversibility. Exergy is based on simultaneous energy and entropy balances and can be defined as the maximum amount of work when a system is brought to thermodynamic equilibrium with environment, in terms of temperature, pressure and composition, through reversible process (Szargut *et al*., 88). Thus, exergy is a measure of the maximum useful level of an energy carrier, as imposed by Second Law.

2.2. Exergy Balance

A linear combination between First and Second Laws of Thermodynamics (energy and entropy balances) for an open system with *m* material streams entering, *n* material streams leaving the system and *p* heat transfers, with no significant changes in kinetic and potential energy leads, in steady state, to:

$$\sum_{i=1}^{m} [\dot{m}_e (\bar{h}_e - T_0 \bar{s}_e)]_i - \sum_{i=1}^{n} [\dot{m}_s (\bar{h}_s - T_0 \bar{s}_s)]_i = \dot{W} + T_0 \dot{S}_g + \sum_{i=1}^{p} \dot{Q}_i (1 - \frac{T_0}{T_i}) \tag{1}$$

The second term of this equation is the maximum amount of available work due to the change from the initial to the final conditions of temperature and pressure. In this term, the quantity $T_0 S_g$ is the lost work (exergy destruction). This loss is due to entropy generation (S_g), caused by process irreversibilities. In the particular case where an entering stream achieves physical equilibrium with environment (T_0, P_0) through reversible processes ($S_g = 0$), the work obtained is the maximum possible and represents the total physical exergy of the stream. For systems in physical equilibrium with environment but with different composition, there is a remaining potential, which can be used to cause changes. The chemical exergy (b_{ch}) is therefore defined as:

$$\bar{b}_{ch} = \sum_{i=1}^{q} x_i (\mu_{0i} - \mu^*_{0i}) \qquad [\text{kJ/kg}] \tag{2}$$

Where x_i is the mole fraction of the component *i* in the system, μ_{0i} is the chemical potential of each component and $\mu_{0i}{}^*$ the chemical potential of the substance (or reference substances) on defined environmental conditions.

2.3. Applications

2.3.1. Industrial use

For energy conversion systems, recent exergy analysis has been used successfully for evaluation and optimization (Hammond, 2007; Kanoglu and Dincer, 2009) and evaluation of environmental impacts (Meyer *et al.*, 2009). The literature presents many cases in different types of industries: sugarcane industry (Pellegrini and Oliveira Jr., 2007; Serra *et al.*, 2008; Kamate and Gangavati, 2009), petrochemicals (Rivero, 2002; Bernardo *et al.*, 2006) and hydrogen production (Orhan *et al.*, 2009). Improvements on process energy efficiencies can lead to the same level of production and services with less consume and depletion of natural resources, matching principles 4,5 and 10. Therefore, exergy, as an auxiliary tool for industrial efficiency optimization also represents a potential tool to achieve sustainable systems. For energy systems, it can be noted the potential of exergy as auxiliary tool for process integration aiming to maximize the use of available process energy streams as highlighted by Principle 10. Some cases in this area can be found in the works of Serra *et al.* (2008) and in a methanol plant (Anantharaman *et al.*, 2006).

2.3.2. Impacts on ecosystems evaluation

In a general way, the larger the exergy of a pollutant, the larger the disturbance caused on environmental equilibrium (Kotas, 95). Being a measure of the difference between a given amount of matter and the environment, the exergy content of an emission represents the maximum potential environmental changes until the matter reaches total equilibrium with the environment. Environmental pollution can be interpreted as any interference in the ecosystem equilibrium through material or energy streams. Such streams cause unbalances, due to differences concerning the environmental parameters (temperature, pressure and composition). Exergy, better than energy, reflects these differences. Further discussions about ecosystem evaluation by exergy can be found in the book of Dincer and Rosen (2007): Exergy: Energy, Environment and Sustainable Development.

2.3.3. Depletion of natural resources evaluation

Resources are materials in disequilibrium with the environment, what leads to an exergy content. This is the useful value of resources, not complete described by energy or mass. Minimize depletion of these natural resources is one of the principles of sustainable engineering (Abraham, 2006). Exergy can be used as an auxiliary tool for sustainability indicators of renewable resources as highlighted by Omer (2008). Regarding Principle 6 (Table 1), a measure is required that is able to account for the quality losses during recycling. These quality losses cannot be measured by mass balances, as the quality degradation cannot be translated by mass measures alone (Amini *et al.*, 2007); exergy is then proposed as a measure of the efficiency of resources use. An interesting concept that arises from this discussion is the thermo-ecological cost, defined as the cumulative consumption of non-renewable exergy due to the production of a particular product (Szargut and Stanek, 2008). Further discussions of how thermodynamics plays a role on economic level and not only on industrial process level can be also found in Ayres (2008) and in Jorgensen (2006).

3. ELCA: Exergetic Life Cycle Assessment

Life Cycle Analysis (LCA) is a holistic approach to cope with environmental performance of a product, based on its entire life cycle, from the raw material acquisition to ultimate disposition. LCA is very useful to identify all material and energy flows consumed and produced during the entire life period of products.

However, one of the limitations of this method is that it cannot evaluate different emissions in regard to environmental effects in the same basis. As exergy can clarify and quantify resource depletion, waste emission and process losses when comparing different substances, there is an increasing number in literature reports of application of the so-called Exergetic Life Cycle Analysis (ELCA) (Cornelissen and Hirs, 2002), as reported for electricity generation (Bakshi and Ukidwe, 2006), hydrogen production (Granovskii et al., 2007) and coal gasification and aluminium production (Ulgiati et al., 2006). There are three major advantages of using exergy in LCA: 1. Provide a common measure of inputs and outputs and allows the estimation of exergetic efficiency that is an indication of potential improvements. 2. Facilitate the comparison between different materials in relation to environmental impacts. 3. Facilitate reporting and monitoring of companies and countries of environmental indicators over time (Ayres *et al*., 1998).

4. Cases

4.1. Efficiency in use of resources in a cogeneration plant
We applied the exergy efficiency concept in a cogeneration plant in a pulp and paper industry (Figure 1). The resources are biomass, black liquor (sub-product of Kraft papermaking) and fossil fuel-based oil. By mean of physical exergy of different streams, two assessment of efficiency can be done: energy basis or utilizable part of energy (exergy). An exergy efficiency, ψ, can be defined as $\psi = E_{out}/E_{in}$, which represents the portion of resources that is delivered in useful form by the products. The energetic efficiency is about 81.55%, while in exergy basis, is 24,90%. Thus, exergy approach clearly reveals that most part of exergy stocked in fuels is not converted, but destroyed by irreversible phenomena taking place in the system. As noted by Rosen et al. (2008), the use of energy as a measure for identifying and measuring the benefits of energy systems can be misleading and confusing, while exergy identify possible efficiency improvements.

4.2. Ammonia process
Figure 3 shows the typical mass and energy flows of an ammonia process, taken form Bakshi and Ukidwe (2006). It reveals the ability of exergy analysis for cope with different materials and energy flows in the same and rational basis. In the first case, different units are used leading to a lack of information about the quality of the resources. In contrast, exergy flows shows the streams in terms of consistent units of exergy. It is also showed the value of the so-called cumulative exergy consumption (436 MW) which considers exergy losses in the process and its supply chain up to natural resources.

5. Concluding Remarks

Alternatives thermodynamic approaches for energy efficiency assessment and depletion of natural resources can provide additional and useful information seeking for sustainable processes. Create engineering solutions beyond current or dominant technologies, improving and innovate to achieve sustainability is one of the principles of sustainable engineering (Abraham, 2006) and exergy analysis (a Second Law based concept) can match this principle as been an auxiliary and non-conventional tool. Matter and energy are always conserved in any process they are used. The real indication of natural resources depletion is related to their useful value (exergy) in conducting process and maintaining suitable conditions of life on Earth.
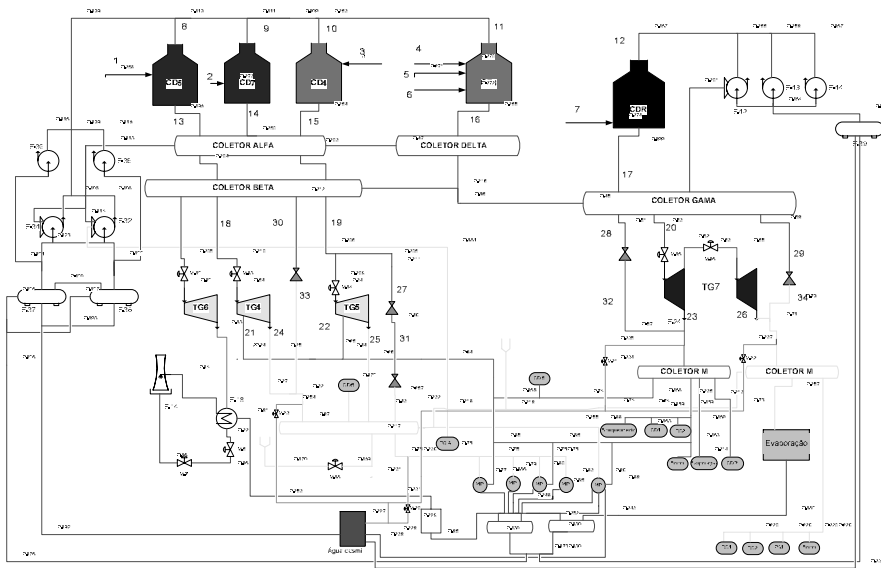
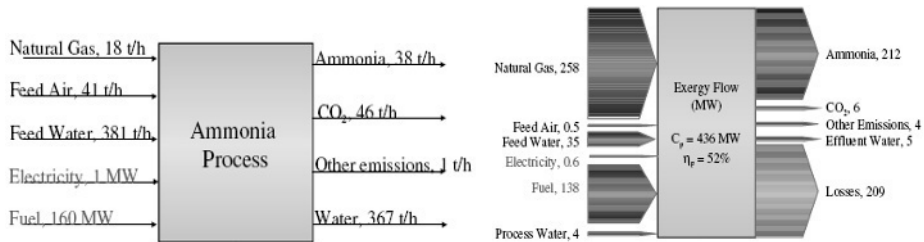**Figure 1: Cogeneration system in a pulp and paper industry.**

Figure 2: Energy and exergy flows form a typical ammonia process. Source: Bakshi and Ukidwe (2006).

References

Abraham, M.A., 2006, Principles of Sustainable Engineering, Sustainability Science and Engineering: Defining principles, M.A. Abraham, Elsevier B.V., Amsterdam, The Netherlands.

Amini, S.H., Remmerswaal, J.A.M., Castro, M.B., Reuter, M.A., 2007, Quantifying the quality loss and resource efficiency of recycling by means of exergy analysis, Journal of Cleaner Production, 15, 907-913.

Anantharaman, R., Abbas, O. S., Gundersen, T., 2006, Energy Level Composite Curves - a new graphical methodology for the integration of energy intensive processes, Applied Thermal Engineering, 26, 1378–1384.

Anastas, P.T., Zimmerman, J.B., 2006, The Twelve Principles of Green Engineering as a Foundation for Sustainability, Sustainability Science and Engineering: Defining principles, M.A. Abraham, Elsevier B.V., Amsterdam, The Netherlands.

Ayres, R.U., 2008, Sustainability economics: Where do we stand?, Ecological Economics, 67, 281-310.

Ayres, R.U., Ayres, L.W., Matinás, K., 1998, EXERGY, WASTE ACCOUNTING AND LIFE-CYCLE ANALYSIS, Energy, 23, 355-363.

Bakshi, B.R., Fiksel, J., 2003, The Quest for Sustainability: Challenges for Process Systems Engineering, AIChE Journal, Vol. 49, No. 6, 1350-1358.

Bakshi, B.R., Ukidwe, N.U., 2006, The Role of Thermodynamics in Life Cycle Assessment of Existing and Emerging Technologies, Proceedings of the 2006 IEEE International Symposium on Electronics and the Environment.

Bernardo, P., Barbieri, G., Drioli, E., 2006, AN EXERGETIC ANALYSIS OF MEMBRANE UNIT OPERATIONS INTEGRATED IN THE ETHYLENE PRODUCTION CYCLE, Chemical Engineering Research and Design, 84, 405–411.

Charpentier, J.C., 2008, Perspective on multiscale methodology for product design and engineering, Computers and Chemical Engineering, In Press.

Cornelissen, R. L., Hirs, G.G., 2002, The value of the exergetic life cycle assessment besides the LCA, Energy Conversion and Management, 43, 1417-1424.

Dincer, I., Rosen, M.A., 2007, Exergy: Energy, Environment and Sustainable Development, Elsevier.

Garcia-Serna, J., Perez-Barrigon, L., Cocero, M.J., 2007, New trends for design towards sustainability in chemical engineering: Green engineering, Chemical Engineering Journal, 133, 7–30.

Granovskii, M., Dincer, I., Rosen, M.A., 2007, Exergetic life cycle assessment of hydrogen production from renewables, Journal of Power Sources, 167, 461–471.

Hammond, G.P., 2007, Industrial energy analysis, thermodynamics and sustainability, Applied Energy, 84, 675–700.

Jorgensen, S.E., 2006, Eco-exergy As Sustainability, WIT Press, Southampton.

Kamate, S.C., Gangavati, P.B., 2009, Exergy analysis of cogeneration power plants in sugar industries, Applied Thermal Engineering, 29, 1187–1194.

Kanoglu, M., Dincer, I., 2009, Performance assessment of cogeneration plants, Energy Conversion and Management, 50, 76–81.

Kotas, T., 1995, The Exergy Method of Thermal Plants Analysis, Melbourne, Krieger.

Meyer, L., Tsatsaronis, G., Buchgeister, J., Schebek, L., 2009, Exergoenvironmental analysis for evaluation of the environmental impact of energy conversion systems, Energy, 34, 75–89.

Ness, B., Urbel-Piirsalu, E., Anderberg, S., Olsson, L., 2007, Categorising tools for sustainability assessment, Ecological Economics, 60, 498 – 508.

Omer, A.M., 2008, Energy, environment and sustainable development, Renewable and Sustainable Energy Reviews, 12, 2265–2300.

Orhan, M. F., Dincer, I., Rosen, M. A., 2009, The oxygen production step of a copper–chlorine thermochemical water decomposition cycle for hydrogen production: Energy and exergy analyses, Chemical Engineering Science, 64, 860 – 869.

Pellegrini, L. F., Oliveira Jr, S., 2007, Exergy analysis of sugarcane bagasse gasification, Energy, 32, 314–327.

Rivero, R., 2002, Application of the Exergy Concept in the Petroleum Refining and Petrochemical Industry, Energy Conversion and Management, 43, 1199-1220.

Rosen, M.A., Dincer, I., Kanoglu, M., 2008, Role of exergy in increasing efficiency and sustainability and reducing environmental impact, Energy Policy, 36, 128–137.

Serra, L.M., Lozano, M., Ramos, J., Ensinas, A.V., Nebra, S.A., 2008, Polygeneration and efficient use of natural resources, Energy, In Press.

Szargut, J., Morris, D., Steward, F., 1988, Exergy Analysis of Thermal, Chemical and Metallurgical Processes, New York, Hemisphere.

Szargut, J., Stanek, W., 2008, Influence of the pro-ecological tax on the market prices of fuels and electricity, Energy, 33, 137–143.

Ulgiati, S., Raugei, M., Bargigli, S., 2006, Overcoming the inadequacy of single-criterion approaches to Life Cycle Assessment, Ecological Modelling, 190, 432–442.

Yi, H., Hau, J.L., Ukidwe, N.U., Bakshi, B.R., 2004, Hierarchical Thermodynamic Metrics for Evaluating the Environmental Sustainability of Industrial Processes, Environmental Progress, Vol.23, No.4, 302-314.

Evaluation of Adsorbed Polyampholyte Layers by Using Quartz Crystal Microbalance

Deusanilde J. Silva[a], Orlando J. Rojas[b], Song Won Park[a], Martin A. Hubbe[b]

[a] *Department of Chemical Engineering - USP Av. Prof. Luciano Gualberto, n. 380, trav. 3, CEP 05508-900, São Paulo - SP, Brazil. deusanilde@gmail.com*
[b] *Department of Forest Biomaterials, NC State University, Raleigh, NC, USA*

Abstract

Viscoelastic properties of layers of polyampholytes adsorbed on charged surfaces were studied by quartz microgravimetry. By applying the Voigt viscoelastic model the effective mass and thickness of layers after adsorption from solution at different salt concentrations were calculated. The obtained results were compared with the Sauerbrey equation, which applies to the case of thin, rigid adsorbed layers. The estimates of mass and thickness from the Voigt model were typically larger, and were more strongly affected by variations in the ionic strength of adsorbing solution. Since the Voigt model uses multiple frequencies and dissipation overtones, it was found that the calculated adsorbed layer mass was closer to the actual values, while the Sauerbrey approach resulted in underestimation. This observation was explained by the fact that adsorbed layers of polyampholytes were soft and highly dissipative. It was noted that the observed changes in dissipation of the adsorbed polyampholyte layers were comparatively large, which suggests a large amount of coupled water.

Keywords: Quartz microgravimetry; Sauerbrey equation; Voigt model; Adsorbed mass; Layer thickness.

1. Introduction

Polyampholytes are natural or synthetic organic polymers that have both, positive and negative groups in their macromolecular structure (Bohidar, 2002). A distinctive feature of these macromolecules is their anti-polyelectrolyte behavior, which involves low solubility in water near a pH value corresponding to net-neutral charge, and the fact that their solubility at such pH increases with increased concentration of salt in solution. This behavior is explained by the balance of electrostatic interactions between positive and negative groups in the molecule. On the other hand, with increasing concentration of low molecular weight electrolytes, the electrostatic attraction between the opposing groups of charges is reduced, resulting in increased expansion and solubility of the polymer in solution. Polyampholytes have found increased utilization in a wide range of applications in medicine, wastewater treatment, oil recovery and papermaking, to name only a few. Adsorption of charged polymers depends on bulk solution environmental conditions, the charge density of the polymer and the surface charge (Rojas, 2002). Evaluation of the polymer adsorbed layer can be carried out by using the Quartz Crystal Microbalance (QCM) (Rodahl and Kasemo, 1996; Höök et al., 2002; Munro and Frank, 2004; Silva, Yamaguchi et al., 2008; Silva, Hubbe et al., 2008). This device enables studies of the dynamics of polymer adsorption processes by measuring the shift in resonance frequency of a quartz crystal electrode with time. The energy dissipation, related to the damping of the sensor oscillation, can be used to study the viscoelastic properties of the adsorbed layer, i.e., its softness or rigidity. Typically, a soft layer is able to more effectively dampen the crystal oscillation. Information from QCM is

relevant to understand the relationship between the structure of the adsorbed layer and the hydration behavior of macromolecules at interfaces. The thickness and the mass of an adsorbed polymer, with its coupled water can be estimated by using Sauerbrey equation. This equation assumes that the amount of adsorbed mass is proportional to the changes in QCM frequencies, which is the case for adsorbed rigid layers (Sauerbrey, 1959). The QTool software of the Quartz Crystal Microbalance with Dissipation (E4 QCM-D unit from Q-Sense, Sweden) was used to evaluate the adsorption of polyampholyte layers from aqueous solution. The changes in frequencies and energy dissipation at different overtones were used to quantify the adsorbed layer properties. Overall, the main objective was to evaluate the adsorbed polyampholyte layers with regard to the adsorbed mass and thickness, under different salt concentration. This was accomplished by using the Sauerbrey equation and the Voigt viscoelastic model.

2. Experimental

2.1. Materials

2.1.1. Polymer

The amphoteric polymer was provided by Harima Chemicals Inc., Japan. It was prepared by random, free-radical polymerization. The cationic monomer was dimethylaminopropilacrylamide, a tertiary amine, the anionic monomer was itaconic acid and the neutral was acrylamide. Molecular weight of the polyampholyte synthesized was 2.93×10^6 Da with cationic to anionic group ratio of 5:4, as measured by NMR (Wang et al., 2002). Aqueous solution (1g/l) viscosity of 2,400 mPa.s (25 ^{o}C) (Wang et al., 2002) and isoelectric point of 7.3 (Silva, Hubbe et al., 2008).

2.1.2. Model surfaces

Cellulose films were developed by following the protocol described by (Gunnars et al., 2002). Gold-coated quartz sensors (for QCM-D measurements) were used as base substrate for the cellulose films. The surfaces obtained by spin coating deposition consisted of flat, uniform and ultrathin films. The roughness of the dry model films were about 5 nm (root-mean-square value) as measured by atomic force microscopy (AFM). Silica substrates bearing different charge density than that of cellulose, was also used for comparison purposes. Gold and silica sensors were provided by Q-Sense, Sweden.

2.2. Density and viscosity

The densities of the polyampholyte solutions for each salt concentration were obtained by measuring the weight of a fixed volume of 5000 μl with 0.0001 g of accuracy (Adventurer SL from Ohouas). Viscosity measurements were performed according to the TAPPI T230 om-89 protocol by using a viscosimeter from Cannon Instrument Company. Both viscosity and density were measured at room temperature, 25 °C. Four replicates were made of each solution to obtain both parameters.

Table 1 - Fixed fluid density and viscosity parameters for Q-Tools input

Parameter	Salt concentration, mM [NaCl]				
	0.1	1	10	100	1000
Fluid density (ρ_f), kg/m3	1002.94	1000.23	1002.31	1005.74	1043.72
Fluid viscosity (η_f), kg/m.s	0.00107	0.00109	0.00100	0.000968	0.00187

2.3. Models

The thickness and the mass of adsorbed polyampholyte were calculated from the QCM-D data by using two approximations, namely, the Sauerbrey rigid model and the Voigt viscoelastic model. For the Voigt model we used *Q-Tools* software to carry out the respective determination, where three harmonics were used for frequency and

dissipation responses under five ionic strength and two substrate conditions. The changes in QCM frequencies were assumed to be proportional to the adsorbed mass, as is the case of adsorbed rigid layers (Sauerbrey, 1959). Alternatively, we used the Voigt model (Voinova et al., 1999) to relate the adsorbed mass and the viscoelasticity of the film. The fundamental frequency fo (oscillating frequency without adsorbed mass) was 4.95 MHz. The sensitivity constant (C) in the Sauerbrey equation has the value 17.7 ng.Hz^{-1}.cm^{-2}. Thus the added or adsorbed mass Δm (ng/cm^2) was calculated from Eq. 1, where n is the overtone number ($n = 1, 3, 5, 7$). If a rigid layer is evenly deposited on one or both sides of the electrodes, the resonant frequency will decrease proportionally to the mass of the adsorbed layer (Garg et al., 2008).

$$\Delta m = -\frac{C.\Delta f}{n} \text{ ; and rearranging to the Sauerbrey equation: } \Delta f_m = -\frac{2 \times f_0^2 \times \Delta m}{A \times \sqrt{\rho_q \mu_q}} \qquad [Eq.1]; [Eq.2]$$

where Δf_m is the measured frequency shift, $\rho_q = 2.648$ g/cm^3 is the density of the quartz, $\mu_q = 2.947 \times 10^{-11}$ dyne/cm^2 is the shear modulus of quartz, and A is the piezo-electrically active area. This equation is not valid if the deposited mass is not rigidly deposited, if it slips on the surface, or if it is not evenly deposited on the electrode surface. The change in the resonant frequency of the QCM crystal also depends on the viscosity and density (η_f, ρ_f) of the gas or liquid medium which is in contact with the crystal according to:

$$\Delta f_{aqueous} = -\frac{n \times f_0^{3/2}}{\sqrt{\pi \rho_q \mu_q}} \left(\rho_f \eta_f \right)^{1/2} \qquad [Eq.3]$$

The viscoelastic properties of the adsorbed polymer layer were evaluated by measuring the energy dissipation of the crystal oscillation, D. Typically, a soft layer is able to quickly dampen the resonator oscillation, while a rigid layer would produce a slower dampening effect. The energy dissipation, or damping, was thus defined by Eq. 4 (Rodahl and Kasemo, 1996), where f is the fundamental resonance frequency and τ is the decay time constant of the oscillation amplitude,

$$D = \left(\pi f \tau \right)^{-1} \qquad [Eq.4]$$

The QCM-D frequency and dissipation data were monitored with time before and after polymer injection on the different substrates at different conditions of ionic strength. Rinsing with buffer solution at the end of the adsorption experiments was typically performed to determine the amount that was irreversibly adsorbed. During the measurements, the QCM liquid chamber was temperature-stabilized to 25 °C and buffer solution was injected at a flow rate of 130µl/min. All experiments with the QCM were repeated at least two times. See the data modeling procedure in Figure 1.
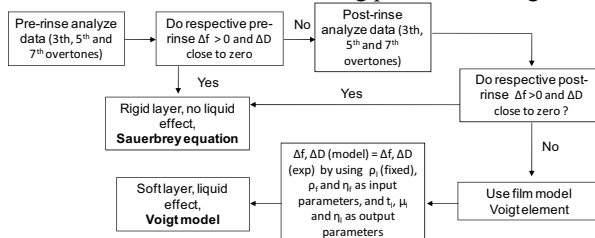

Figure 1 – Flowchart of QCM-D data modeling, subscript l is for "layer" and f is for "fluid". After: (Munro and Frank, 2004).

The Figure 2 shows the Voigt model where the polymer layer is subjected to an oscillating shear stress σ and behaves like a Voigt element with shear viscosity η_l as a dashpot and shear modulus μ_l as a spring. (Voinova et al., 1999; Tammelin et al., 2004).
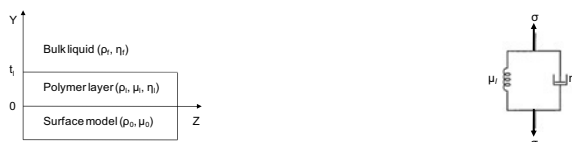

Figure 2 – Geometry of the surface model covered by a viscoelastic mono-layer (left) and a schematic depiction of Voigt viscoelastic element (right).

The derivation of Voigt model (Voinova et al., 1999) as described by (Vogt et al., 2004) is

$$\Delta f \approx \frac{t_l \rho_l f}{2\pi \rho_0 t_0}\left(1+\frac{2t_l^2 \chi}{3\delta^2(1+\chi^2)}\right) \; ; \; \Delta D \approx \frac{2t_l^3 \rho_l f}{3\pi f_0 \rho_0 t_0}\frac{1}{\delta^2(1+\chi^2)} \; ; \; \chi = \frac{\mu_l}{\eta_l f} ; \; \delta = \sqrt{\frac{2\eta_l}{\rho_l f}} \qquad \text{[Eqs.5-8]}$$

where χ is the ratio of the storage modulus μ_l and the loss modulus η_l. The length δ is the viscous penetration depth.

3. Results and discussion

3.1. Suitability of the viscoelastic model

The dissipation values of the polyampholyte for both silica and cellulose surfaces for all salt concentrations were larger than zero for both pre-rinsing and post-rinsing steps (Figure 3). Substantial deviations from Sauerbrey equation could be expected to occur (Rodahl and Kasemo, 1996; Höök et al., 2002), and a viscoelastic model is therefore recommended (Rodahl and Kasemo,1996; Voinova et al., 1999).
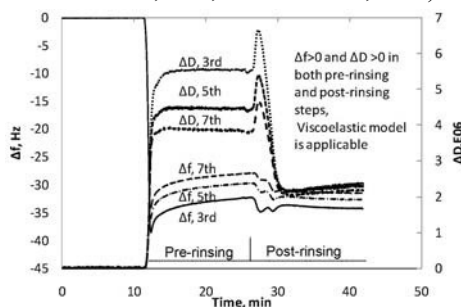

Figure 1 – Changes in frequency and dissipation for polyampholyte adsorbed layer on silica surface at pH 4.3 and 10 mM of salt concentration.

3.2. Experimental and fitted results

The Voigt model was used to fit the thickness and the amount of adsorbed polymer mass. Figure 4 shows the experimental and fitted results for the three harmonics studied (3rd, 5th and 7th), and the comparison between the Sauerbrey and fitted thickness.

A viscoelastic model to evaluate the thickness and the polymer adsorbed mass was more applicable than the Sauerbrey model for rigid films. For example, the thickness of the absorbed polymer layer estimated by using Voigt model was 5 times higher than that estimated by using the Sauerbrey equation, as shown in Figure 4 right. Figure 5 shows the adsorbed mass and the thickness of the polyampholyte layers for silica (left) and cellulose (right) surfaces at five salt concentrations. Higher values of adsorbed mass and thickness were observed at intermediate salt concentrations. It is noted that the observed

changes of dissipation of the adsorbed polyampholyte layers were comparatively large, which suggests a large amount of coupled water.
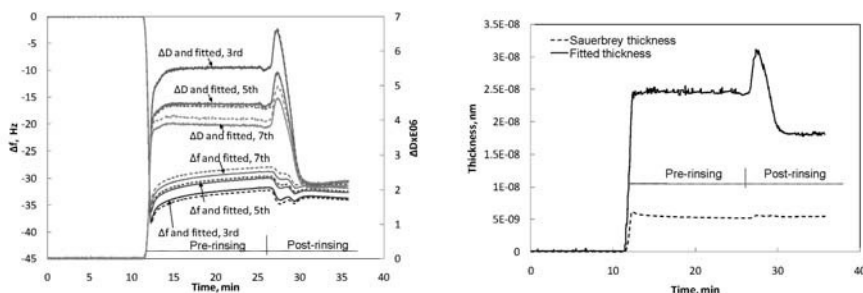

Figure 2 – Comparison between experimental and fitted results by using Voigt model (left), and thickness results by using Sauerbrey Equation and Voigt Model (right). Both graphics were for polyampholyte layer adsorbed on silica surface at pH 4.3 and 10 mM [NaCl].

At low electrolyte concentration, the polyelectrolyte adsorption can be attributed to an ion-exchange mechanism that happens in the electrostatic double layer (Rojas et al., 2001; Rojas, 2002). This mechanism is thermodynamically favored due to the net gain of entropy due to release of counter-ions at the interphase. At high electrolyte concentration, the electrostatic interaction between the polyelectrolyte and the charged surface is reduced. Higher fitted values for polymer adsorbed amount and thickness at the highest salt concentration (1000 mM) on the cellulose surfaces were not expected (Figure 5 B). These results can be explained by the large variability of the results recorded in QCM trials for this salt concentration condition that was computed in the fitting analysis.
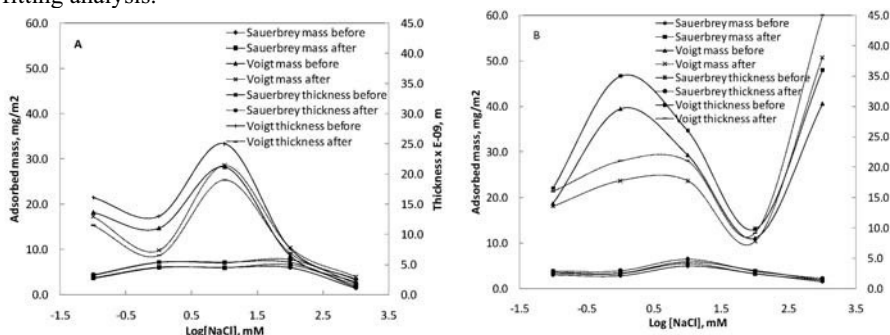

Figure 3 - Results of adsorbed amount of mass and thickness of the polyampholyte layer for silica (A) and cellulose (B) model surfaces.

Surface and polymer charge densities, and ionic strength of adsorbing solution, are factors that affect the interactions forces at the solid - liquid interface (Rojas, 2002). The conformation of these polymers in bulk solution also affects the characteristics of the adsorbed layer, its rigidity and viscoelasticity. Although we expected that these parameters increase with salt concentration (as well as the dynamic light scattering and turbidity, not presented here), we show that at higher salt concentration the electrostatic screening do not allow effective interactions between the polymer and the surface.

4. Conclusion

In the present work we evaluated the properties of adsorbed polyampholyte layers under different salt concentrations. Calculation were performed following the Sauerbrey Equation and a viscoelastic model, for both silica and cellulose surfaces. According to the results, the following conclusions can be drawn: (1) Larger polymer adsorbed mass and thickness were obtained with the Voigt model; (2) The Voigt model showed larger effects of the variations from changes in the ionic strength of adsorbing solution that otherwise is not accounted by Sauerbrey approach; (3) Since the Voigt model uses multiple frequency and dissipation overtones, it was found that the calculated adsorbed layer mass is closer to the actual mass while the Sauerbrey approach underestimates the respective values. Such underestimation is explained by the fact that the adsorbed polyampholyte layers are soft and dissipative.

References

Bohidar, H. B., Ed. (2002). <u>Characterization of polyelectrolytes by dynamic laser light scattering</u>. Handbook of Polyelectrolytes and their Applications. California, USA, American Scientific Publishers.

Garg, A., Heflin, J. R., Gibson, H. W., Davis, R. M. (2008). "Study of film structure and adsorption kinetics of polyelectrolyte multilayer fims: effects of pH and polymer concentration." <u>Langmuir</u> **24**: 10887-10894.

Gunnars, S., Wågberg, L., Cohen Stuart, M.A. (2002). "Model films of cellulose: I. Method development and initial results." <u>Cellulose</u> **9**: 239–249.

Höök, F., Vörös, J., Rodahl, M., Kurrat, R., Böni, P., Ramsden, J.J., Textor, M., Spencer, N.D., Tengvall, P., Gold, J., Kasemo, B. (2002). "A comparative study of protein adsorption on titanium oxide surfaces using in situ ellipsometry, optical waveguide lightmode spectroscopy, and quartz crystal microbalance/dissipation." <u>Colloids and Surfaces B: Biointerfaces</u> **24**: 155-170.

Munro, J. C., Frank, C. W. (2004). "Polyacrylamide Adsorption from Aqueous Solutions on Gold and Silver Surfaces Monitored by the Quartz Crystal Microbalance." <u>Macromolecules</u> **37**(3): 925-938.

Rodahl, M., Kasemo, B. (1996). "Frequency and dissipation-factor responses to localized liquid deposits on a QCM electrode." <u>Sensors and Actuators A-Physical</u> **54**(1-3): 448-456.

Rojas, O. J. (2002). Adsorption of polyelectrolytes on mica <u>Encyclopedia of Surface and Colloid Science</u>. A. Hubbard. New York, Marcel Dekker**:** 517-535.

Rojas, O. J., Neuman, R.D., Claesson, P.M. (2001). "Desorption of low-charge-density polyelectrolyte adlayers in aqueous sodium n-dodecyl sulfate solution." <u>Journal of Colloid and Interface Science</u> **237**: 104-111.

Sauerbrey, G. Z. (1959). "The use of quartz oscillators for weighing thin layers and for microweighing." <u>Zeitschrift Fur Physik</u> **155**(2): 206-222.

Silva, D. J., Hubbe, M. A., Park, S. W., Rojas, O. J. (2008). Adsorption and viscoelastic properties of polyampholytes monitored by QCM-D. <u>V CONGRESO IBEROAMERICANO DE INVESTIGACION EN CELULOSA Y PAPEL 2008</u>. Mexico.

Silva, D. J., Yamaguchi, T., Song, J., Park, S.W., Hubbe, M., Rojas, O. (2008). <u>Swelling and water-holding ability of adsorbed polyampholytes</u>. 235th ACS National Meeting an Exposition, New Orleans, LA.

Tammelin, T., Merta, J., Johansson, L.-S., Stenius, P. (2004). "Viscoelastic properties of cationic starch adsorbed on quartz studied by QCM-D." <u>Langmuir</u> **20**(25): 10900-10909.

Vogt, B. D., Lin, E. K., Wu, W.-L., White, C. C. (2004). "Effect of Film Thickness on the Validity of the Sauerbrey Equation for Hydrated Polyelectrolyte Films." <u>J. Phys. Chem. B</u> **108**: 12685-12690.

Voinova, M. V., Rodahl, M., Jonson, M., Kasemo, B. (1999). "Viscoelastic acoustic response of layered polymer films at fluid-solid interfaces: Continuum mechanics approach." <u>Physica Scripta</u> **59**(5): 391-396.

Wang, Y., Hubbe, M., Sezaki, T., Wang, X., Rojas, O., Argyropoulos, D. (2002). "The role of polyampholyte charge density on its interactions with cellulose." <u>Nordic Pulp & Paper Research Journal,</u> **21**(5): 158-165.

Optimization of Scaffolds in Alginate for Biofabrication by Genetic Algorithms

Rodrigo Rezende[a], Mylene Rezende[a], Paulo Bártolo[b], Ausenda Mendes[b],
Rubens Maciel Filho[a]

[a]*University of Campinas, P.O. Box 6066, Campinas, 13083-970, Brazil*
[b] *Centre for Rapid and Sustainable Product Development, Polytechnic Institute of Leiria, Portugal*

Abstract

With an increasing in the rate of transplants due to damaged or affected tissues or organs by accidents or diseases and also by the aging of the population in many countries as Brazil, have motivated the research of some novel and alternative ways focused on restoring and replacing tissues. Biofabrication by means of Rapid Prototyping techniques can help in the fashioning and final production of scaffolds devoted to support and stimulate the growth of new tissues. For soft tissues, a biomaterial known as Alginate has been studied and used as raw-material for scaffolds fabrication. A scaffold must guarantee good strength and stiffness at the same time the material degrades gradually. In this work, a single mathematical model experimentally obtained that describes an interesting mechanical behavior of the degradation of alginated-scaffolds is developed. The optimization process scheme using Genetic Algorithms to maximize the elastic modulus and therefore to aid the design of scaffolds in alginate is proposed. The optimization is very welcome to tissue engineering and Biofabrication.

Keywords: Genetic Algorithms (GAs), Scaffolds, Biofabrication, Tissue Engineering, Alginate.

1. Introduction

A Brazil's recent picture of the 2008 waiting list for transplants, according to the Ministry of Health, reveals that there are more than 68,000 candidates waiting the availability of organs to therefore undergo a transplant. Based on those data, alternative methods of tissue and organ recovering have been studied and many developments have been proposed and applied successfully in the Tissue Engineering. Tissue Engineering as an interdisciplinary field combines the use of living cells with either natural or synthetic extra-cellular structures (scaffolds) to develop body parts or devices that will enable the restoration, maintenance or enhancement of living tissue and organs (Rezende et al., 2007a). Usually, the scaffolds have high porosity (macro-porosity), appropriate surface morphology (micro-porosity), large surface area, suitable pore size and highly connected pore structure. They must also be biocompatible and biodegradable. A variety of biodegradable and biocompatible hydrogels, as the alginate, have been used for Tissue Engineering. The alginate is one of the most popular materials due to its relatively low cost, natural origin and easy handling besides other physical properties advantages. Biofabrication is a new class of Rapid Prototyping that

represents a group of non-conventional techniques with great potential to produce scaffolds with customised external shape and predefined internal morphology (Leong et al., 2003). These techniques also allow controlling both pore size and distribution. A scaffold must own very dynamical and adaptive characteristics in order to be implanted and to take its main roles which are to carry the stem live cells inside it, to back the growth of these cells and besides this to biodegradate appropriately since the minimum material should remain after the tissue is reconstructed. In this sense, it is fundamental to be aware of the mechanical and chemical properties since the scaffold must guarantee good strength and stiffness at the same time the material degrades gradually. To know how the mechanical behavior of the scaffold will be, some time later, is the keyword. And the understanding about the match between biodegradation and young modulus is mandatory. The present and future of biomedical materials development requires this degree of control prediction in the design, synthesis, and function of next-generation materials. A prediction job is possible and it has already been used so that the scaffold state can be forecasted before its fabrication and, as a good alternative, to know how and how much alginate should be used. Other future analyses can be around the best geometry to be adopted during rapid prototyping technique actuation. This paper presents a single mathematical model experimentally obtained that describes an interesting mechanical behavior of the degradation of alginated-scaffolds. The deal of this work is the optimization of scaffolds in alginate making use of Genetic Algorithms (GAs). GAs represent a class of stochastic optimization procedures based on natural systems according to Darwin's observations. GAs have been successfully applied to a range of problems and have characteristics of easiness of implementation and capability of escaping local optimal solution. The objective of GA is to find out the best values of alginate amount for scaffold fabrication that maximize the elastic modulus.

2. Optimization Problem Formulation

The aim of the optimization is to determine optimal features for the fabrication of optimized alginate scaffolds for Tissue Engineering. The optimization objective in this work is to find out optimal values of alginate composition and initial porosity in order to fabricate scaffolds with, at a pre-determined time, a high mechanical behavior (elastic modulus), since those parameters are two of the most important ones and they need to be well combined in order to reach the best value for the elastic modulus. Then, the optimization problem can be written as:

$$\underset{[\alpha,\phi_0]}{\text{Maximize}} \quad E(\phi_0, \alpha, t)$$

$$\text{Subject to:} \quad \begin{aligned} 1\% &\leq \alpha \leq 8\% \\ 30\% &\leq \phi_0 \leq 80\% \end{aligned} \tag{1}$$

where E is the elastic modulus (shear effects are not considered), α is the alginate composition and ϕ_0 is the initial porosity.

The variation of the mechanical properties of alginate is due to degradation and porosity changes. The degradation of alginate structures was determined through the analysis of the shrinkage variation along time as shown in Figure 1 (Rezende et al., 2007a):
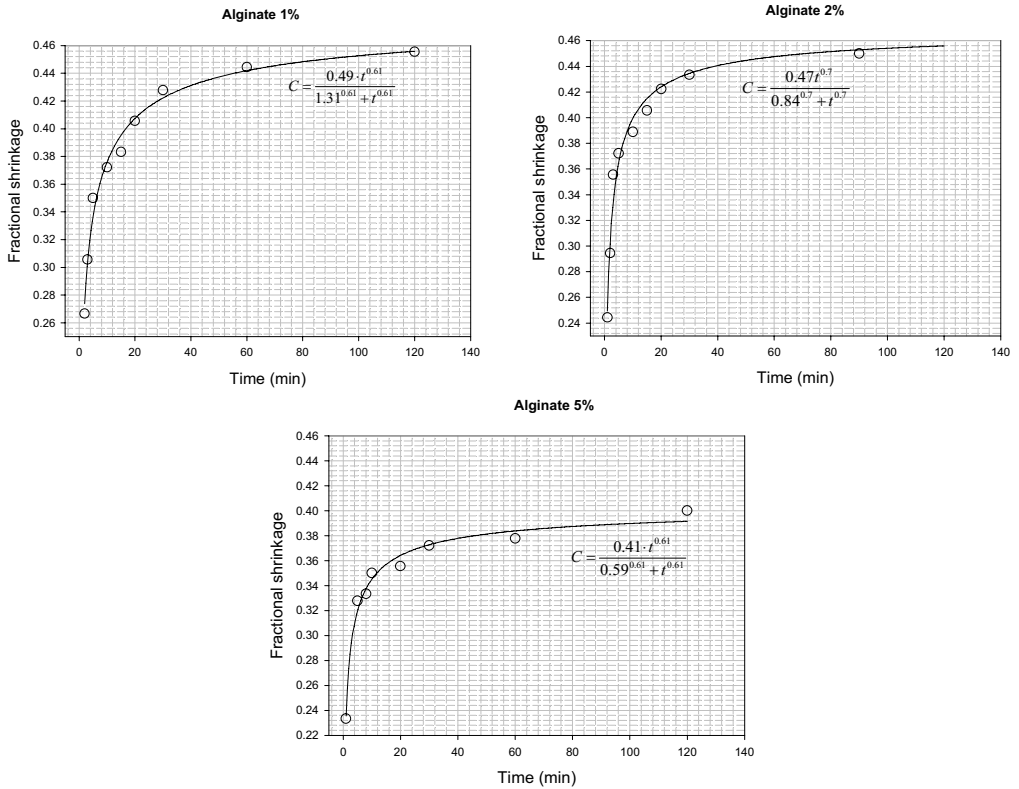
Figure 1 – Fractional shrinkage along time for different alginate compositions (1%, 2%, 5%).

There are various studies modeling tissue scaffolds behaviors in the literature (Nair et al., 2008). In the present case, a sigmoidal model of three parameters represents the shrinkage process that is given by the following equation:

$$C(\alpha, t) = \frac{\varsigma(\alpha) \cdot t^{\vartheta(\alpha)}}{\lambda^{\vartheta(\alpha)} + t^{\vartheta(\alpha)}}$$

(2)

where t is the time and $\varsigma, \vartheta, \lambda$ are variables that depend on the alginate composition (α). Porosity at each time is also a function of alginate composition and shrinkage:

$$\phi(\phi_0, \alpha, t) = \phi_0 + \zeta(\phi_0, \alpha) \cdot C(\phi_0, \alpha, t) + \psi(\phi_0, \alpha) \cdot C^2(\phi_0, \alpha, t)$$

(3)

where ζ, ψ are constants depending on alginate composition and C is the shrinkage.

The dependence between the elastic modulus and porosity for different alginate compositions is given by the following equation:

$$E(\phi_0, \alpha, t) = E_0(\phi_0, \alpha) + k_1(\phi_0, \alpha) \cdot \phi(\phi_0, \alpha, t) + k_2(\phi_0, \alpha) \cdot \phi(\phi_0, \alpha, t)^2 + k_3(\phi_0, \alpha) \cdot \phi(\phi_0, \alpha, t)^3$$

(4)

with E_0 being the initial elastic modulus, k_1, k_2, k_3 constants dependent on both the alginate composition and the initial porosity and ϕ the final porosity of the scaffold.

The optimization problem presented on Equation 1 is a single constrained optimization problem. In this paper, two cases of constraints are considered: constraints at shrinkage and final porosity: 1) shrinkage higher than 25% and 2) final porosity higher than 80%.

In order to solve the constrained optimization problem, a constraint handling method based on the penalty function approach was used, not requiring any penalty parameter (Rezende et al., 2007b). In this case, the expression of the fitness function for a minimization problem, where infeasible solutions are compared based only on their constraint violation, is given by the Equation 5:

$$F(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } g_j(\mathbf{x}) \geq 0 \quad \forall j=1,2,\dots,nc \\ f_{\max} + \sum_{j=1}^{m} \langle g_j(\mathbf{x}) \rangle & \text{otherwise} \end{cases}$$

(5)

where $f_{\max}$ is the objective function value of the worst feasible solution in the population.

3. Optimization of Scaffolds using GAs

The Genetic Algorithms approach starts with a random population of chromosomes that are a set of solutions for the optimization problem. Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration. Usually, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

The Genetic Algorithm used in this research work to solve the scaffolds optimization problem is a Fortran binary code. The employed genetic operators are the tournament selection, the uniform crossover, the creep and the jump mutation. Niching and elitism are also employed. The input parameters, chosen by a trial and error method, are indicated in Table 1.

Table 1 - The GA input parameters.

GA input parameters	Value
Population size per generation	50
Maximum number of generations	30
Crossover probability	0.60
Jump mutation probability	0.077
Creep mutation probability	0.077
Initial random number seed for the GA run	-1000

The following results of the scaffolds optimization using Genetic Algorithms for constrained problem considering two cases of constraints are presented.

3.1. Constraint 1: Shrinkage > 25%

In this case, the objective of the optimization problem is to maximize the elastic modulus subject to a shrinkage higher than 25% and 2) final porosity higher than 80%. The results obtained for this case are shown in Table 2:

Table 2 – Optimization results for the constrained problem (shrinkage > 25%)

Optimization Variables	Initial Alginate composition (%)	7.06
	Initial Porosity (%)	30.00
Objective Function	Elastic modulus (KPa)	17.52
Constraint	Shrinkage (%)	25.22
Output Variable	Final Porosity (%)	70.97

Figure 2 shows the evolution of the objective function along all the generations. Figure 3 presents the obtained profile of the objective function from the best combination of the optimization variables at each generation. The output variables (shrinkage and final porosity) dependent on the best combination are also indicated.
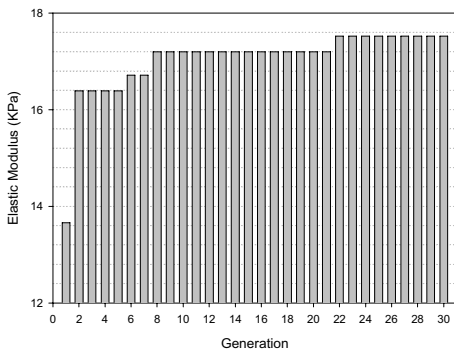

Figure 2 – Evolution of the elastic modulus along all the generations (shrinkage > 25%).
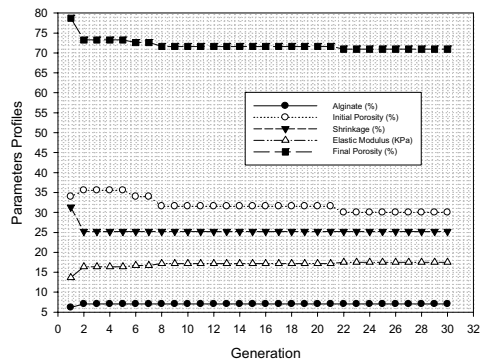
Figure 3 – Profiles of the objective function, shrinkage and final porosity obtained with the best values of the optimization variables at each generation (shrinkage > 25%).

3.2. Constraint 2: Final porosity > 80%

In this case, the objective of the optimization problem is to maximize the elastic modulus subject to a final porosity higher than 80%.

Results obtained for this case are shown in Table 3:

Table 3 – Optimization results for the constrained problem (final porosity > 80%).

Optimization Variables	Initial Alginate composition (%)	5.79
	Initial Porosity (%)	32.38
Objective Function	Elastic modulus (KPa)	12.99
Output Variable	Final Porosity (%)	80.01
Constraint	Shrinkage (%)	33.29

Figure 4 shows the evolution of the objective function along all the generations. Figure 5 presents the obtained profile of the objective function from the best combination of the optimization variables at each generation. The output variables (shrinkage and final porosity) dependent on the best combination are also indicated.
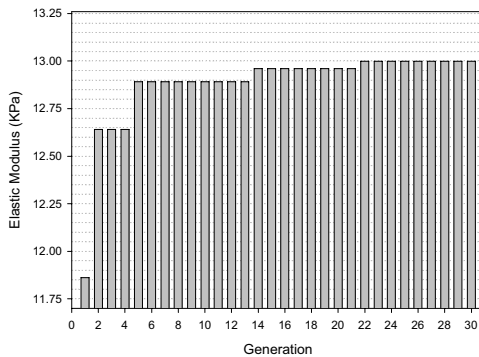
Figure 4 – Evolution of the elastic modulus along all the generations (final porosity > 80%).
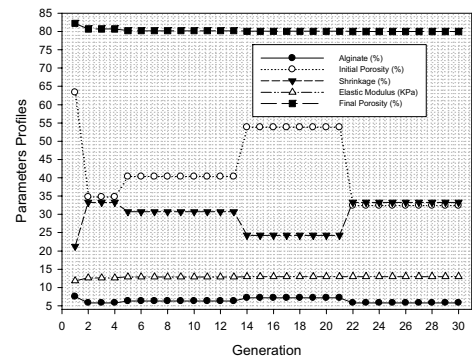
Figure 5 – Profiles of the objective function, shrinkage and final porosity obtained with the best values of the optimization variables at each generation (final porosity > 80%).

Observing the two cases presented, it can be seen that the constraints were respected. In spite of the obtained value of young modulus, the order of this value is compatible with results found in the literature (Khalil, S.E.D., 2005). Throughout this study is incipient and new boundaries and adjustments need to be done. One of the main characteristics of the alginate-scaffold is to degradate along the time where the growing of final porosity indicates this phenomenon. The Initial Alginate composition (α) is the real range that has been used experimentally.

4. Conclusions

This research employs Genetic Algorithms to optimize the mechanical behavior of alginate scaffolds for Biofabrication. The mathematical model was experimentally obtained and the best values for both alginate composition and initial porosity of the scaffold allowing the constrained maximization of the elastic modulus were determined through the optimization by Genetic Algorithms.

References

R.A. Rezende, P.J. Bartolo, A. Mendes, H. Almeida, R. Maciel Filho, 2007a, Experimental Characterisation of the Alginate Gelation Process for Rapid Prototyping. The eighth International Conference on Chemical & Process Engineering, v. 11, p. 509-514.

K.F. Leong, C.M.Cheah, C-K.Chua, 2003, Solid freeform fabrication of three-dimensional scaffolds for engineering replacement tissues and organs. Biomaterials, 24, 2363-2378.

Nair, K., Yan, K. and Sun, W. 2008. A computational modeling approach for the characterization of mechanical properties of 3D alginate tissue scaffolds, Journal of Applied Biomaterials and Biomechanics, 6: 35 – 46.

M.C.A.F. Rezende, R.A. Rezende, A. Mendes, P.J. Bartolo, A.C. Costa, R. Maciel Filho, 2007b, Evolutionary Optimization for Soft Tissue Engineering. Advanced Research in Virtual and Rapid Manufacturing, v.3 , p. 81-85.

Khalil, S.E.D. 2005. Deposition and Structural Formation of 3D Alginate Tissue Scaffolds. Drexel University. (PhD Thesis).

Risk Analysis and Environmental Impact Analysis in a Chemical Processing Facility

Maristhela P. de A. Marin[a], Elias Basile Tambourgi[b]

[a]*Departmentof Chemical Engineering, Centro Universitério da FEI, Av. Humberto de A. C. Branco 3972, São Bernardo do Campo 09850-901, Brasil*
[b]*Faculdade de Engenharia Química,UNICAMP, Av. Albert Einstein 500, Campinas 13083-852, Brasil*

Abstract

The present work provides a method of risk analysis in chemical processing facilities, which consists primarily of the following steps: description and study of the system, hazard analysis and identification, identification of the accidental scenarios, calculation of the consequences of the scenarios and risk characterization. To illustrate the application of the method, an uranium hexafluoride production facility was selected. The main hazards identified were: fire, explosion, spill of toxic chemical substances and ground contamination due to problems in the waterproof blanket of the waste contention bund. Some cases of environmental impact due to uranium hexafluoride (UF_6) release were quantitatively analyzed for different scenarios, considering two categories of atmospheric stability: Pasquill D and Pasquill F. It was found that the impact which reached a great external area around the plant corresponded to the condition of atmospheric stability F.

Keywords: Risk Analysis, Environmental Impact, Consequences Analysis.

1. Introduction

The industrial accidents in the past few years, particularly in the 80's, combined with environmental problems, contributed significantly to gather the attention of government authorities, industry and society, to seek means for the prevention of such episodes that damage people's safety and the quality of the environment.

Currently, there is a strong trend for quantitative analysis of risks, which is considered to be a useful tool in identifying hazards, to be held more frequently on the parts of industrial systems considered most critical with risks of accidents with environmental consequences.

The risk analysis consists on the systematic examination of industry facilities in order to identify hazards and quantify the consequences of an accidental event.

The method of risk analysis proposed in this work is similar to that proposed in World Bank (1985), specifically for chemical processing plants and applicable, with some changes, in nuclear facilities. It was applied to a uranium hexafluoride (UF_6) production plant. In the present work the main hazards were identified and the consequences of accidental release of UF_6 were quantitatively analyzed and the calculations were made using the software Conv_aci (Molnary, 1993).

2. Method

The plant is divided into functional units and for each one the systems and the components of interest are identified. Each component is analyzed individually and therefore determined its possible modes of failure. The inventory of hazardous materials are selected and grouped, being analyzed quantitatively only those with the largest inventories in terms of toxicity, flammability, and quantities. The hazards are identified, the cases of release are classified and the source terms are determined, following the evaluation of consequences. If probability data is known, one can estimate the frequency of undesirable events. Finally, mitigating risks measures are proposed.

An historical analysis of accidents at similar facilities is considered a part of risks analysis. This research aims to identify the hazards in industrial activities, determine the common causes of accidents, raise historical trends and probabilities of occurrence of accidents, assess the results of calculation models simulation and to assist in developing the event trees.

2.1. Case study

The plant chosen for the consequence analysis was Kerr Mc Gee Sequoyah Hexafluoride Plant. This installation purified the concentrated uranium, yellowcake, through the process of extraction by solvent for subsequent conversion into uranium hexafluoride (UF_6). The report (NRC, 1975) describes both plant and chemical processes in detail which allows their analysis. For the application of the method, the plant was divided into four functional units*:*

• Area I: UF_6 production
• Area II: Elemental fluorine production
• Area III: Chemicals storage
• Area IV: Waste contention basins

For all areas components, modes of failure and inventory of materials were identified.

2.1.1. Theoretical background

The assessment of consequences of toxic substances releases requires knowledge of their physicochemical properties and their toxicity, the adoption of a model of atmospheric dispersion and the establishment of criteria for the classification of undesirable events. Although the UF_6 is a stable compound, its manipulation is complex, since the substance is highly reactive with water, ether and alcohol, forming stable products. The reaction of UF_6 with water is highly exothermic and its hydrolysis products are uranyl fluoride (UO_2F_2) and hydrofluoric acid (HF) according to the reaction:

$$UF_{6(g)} + 2H_2O_{(g)} \rightarrow UO_2F_{2\ (s)} + 4HF_{(g)} \tag{1}$$

Of its products of hydrolysis, the HF presents only chemical toxicity while UO_2F_2 has chemical and radiological toxicity. The uranium chemical toxicity surpasses the radiological at any enrichment level (NRC, 1991).

In order to analyze the consequences of UF_6 release, it was considered that the release occurs at ground level, that the individual exposed to the effects of the released substance remains in the center line of the plume during the dispersion and that the particles are highly respirable and transportable. The incorporation of uranium via inhalation is calculated using the expression (NRC, 1986):

$$IU=(\frac{\chi}{Q}).Mu.BR.f_d \tag{2}$$

Where:

IU is the mass of radioactive substance incorporated by inhalation (mg); $\frac{\chi}{Q}$ represents

factor of atmospherical dispersion depending on the distance (s/m^3); *BR* is the breathing rate of an average individual, as $2,66 \times 10^{-4}$ m^3/s; *Mu* is the mass of released uranium (mg); *fd* represents factor of dry deposition correction.

It is considered that the particles resulting from the release of UF$_6$ have about 10^{-6} m of diameter, and therefore, are fully inhaled.

Concentration of HF:
The concentration of HF (C_{HF}) in air is given by the expression (NRC, 1986):

$$C_{HF}= (\frac{MHF}{t}) (\frac{\chi}{Q}) \tag{3}$$

Where:
MHF is the mass of HF formed in the hydrolysis of released UF$_6$; *t* represents time of release (s)

Criteria for classification of the event for the substance release:
To classify the event of accidental release of UF$_6$, two different criteria were used. The first one, proposed by Hanamann et al. (1995) is a quantitative method to classify the impacts on human health shown in Table 1. The groups are defined in terms of tolerance limits on exposure to soluble uranium (*Iu*) and the concentration of hydrofluoric acid gas (C_{HF}), as shown in Table 2. The following indices for the tolerance limits are: *Emergency Response Planning Guideline* (ERPG), *Threshold Limit Value - Short Term Exposure Limit* (TLV-STEL), *Threshold Limit Value - Ceiling* (TLV-C) and *Immediately Dangerous to Life or Health* (IDLH).

Table 1: Classification of the event according to the toxicological consequence

Category	Toxicological Consequence
Despicable	1. Exposition of public inferior than GROUP 3
	2. Exposition on property limits inferior than GROUP 2
	3. Exposition on event location inferior than GROUP 1
Marginal	1. Exposition of public greater than GROUP 3
	2. Exposition on property limits greater than GROUP 2
	3. Exposition on event location greater than GROUP 1
Critical	1. Exposition of public greater than GROUP 2
	2. Exposition on property limits greater than GROUP 21
Catastrophic	1. Exposition of public greater than GROUP 1

Table 2: Values of tolerance limits that define the groups

	Soluble Uranium (*Iu*)	Concentration of HF (C_{HF})
GROUP 3	TLV-STEL: 0,6 mg/m^3	ERPG 1: 4,1 mg/m^3
GROUP 2	TLV-C: 1,0 mg/m^3	ERPG 2: 16,4 mg/m^3
GROUP 1	IDLH: 1,0 mg (incorporation)	ERPG 3: 25 mg/m^3

The second criterion used for classification of UF_6 release is proposed in the report NUREG 1391 (NRC, 1991). In this criterion the event is classified as abnormal or accident due on the individual is exposure to soluble uranium (Iu) and to the concentration of HF (C_{HF}) as shown in Table 3, referenced on the IDLH.

Table 3: Criteria for classification of the event as abnormal or accident

Classification	C_{HF} (mg/m^3)	Iu (mg)
Abnormal Event	< 25 (1800/t)^2	2<Iu<10
Accident	> 25 (1800/t)^2	Iu ≥ 10

3. Results and Discussion

3.1. Major undesirable events

Depending on the method application, the following hazards were identified: fire in the area of extraction by solvent, and UF_6 release from the area of fluorination and collection on Area I; explosion in the fluoride cells in Area II; rupture of substance storage tanks, as anhydrous hydrofluoric acid (HF), anhydrous ammonia (NH_3) and nitric acid (HNO_3), which are used in the process; failure on the sealing blanket of the waste contention basins in Area IV.

This study quantitatively analyzed only the case of UF_6 release and consequently release of HF gas and solid UO_2F_2 as depicted in equation (1).

3.2. UF_6 release

The event analyzed is characterized by the release of liquid UF_6 from the product withdrawal area. It is considered that during the transfer operation of liquid UF_6 from the primary cold trap to the storage cylinder, a rupture of a transport pipeline occurs. It is admitted that all of the liquid UF_6 contained in the cold trap (9492 kg) was released to the process building during 900 seconds (15 min). As a result of the release, some of the product will be solidified and some vaporized. In the plant operating conditions, the vaporized fraction of liquid UF_6 is 49% (Siman-Tov, 1984).

Three scenarios (C) were studied:
• C1: Due to a failure in the ventilation system, the entire amount of UF_6 vaporized (4651 kg, equivalent to 3139 kg of uranium and 1057 kg of HF) is released into the atmosphere, at ground level, during 900 s (15 min).
• C2: The ventilation system failed, but of the total amount of UF_6 vaporized (4651 kg), at ground level, only a portion of the product that reacts with air is released. It is admitted that 80% of HF (846 kg) and 50% of uranium (1560 kg) are released, the rest being confined within the facility.
• C3: Credit is given to the exhaust system, and it is assumed that 90% of the UF_6 vaporized is retained by the filters and gas scrubbers. Therefore, at ground level, it is released to air 314 kilograms of uranium and 106 kg of HF.

According to the distance from the release point, the incorporation of uranium (Iu) and the concentration of HF (C_{HF}) were calculated, considering two classes of atmospheric stability: Pasquill D (neutral atmospheric stability with wind speed equal to 3m/s) and Pasquill F (stable atmospheric stability with wind speed equal to 1m/s) using the software Conv_aci. For event classification, the following areas were defined: Area 1: internal dependencies of the building, where the health impacts of workers of the plant are analyzed; Area 2: within the site, that is, the area bounded by the limits of building

and by the facility property limits, where the health impacts of workers of the plant are analyzed, Area 3: area outside the limits of the facility, where the health impacts on local individuals are analyzed.

It was considered that the limit of the property is located 1175 m from the point of release. Table 4 shows the distance in meters, from which C_{HF} becomes less than 35 mg/m^3, the tolerance limit given by the IDLH for 900 second of leaking and the distance from which Iu is less than 10 mg.

Table 4: Results of UF$_6$ release

	Distance (mts) for $C_{HF}{\leq}35mg/m^3$		Distance (mts) for $Iu{\leq}10mg$	
	Class D	Class F	Class D	Class F
C1	1600	7600	1400	1500
C2	Property Limits	6700	Property Limits	1300
C3	350	1600	500	960

Scenario C1

According to the criterion 1, the toxicological consequences are classified catastrophic for both class D and class F of atmospheric stability, as C_{HF} and Iu to the individual of the public (Area 3) surpass the values of GROUP 1, as indicated on both Tables 1 and 2. Considering the classification by the criterion 2, it is observed that for the Pasquill D stability class, both C_{HF} and Iu are below the reference levels from 1600 m and 1400 m from the release point, which classifies the case as an accident. Considering the atmospheric stability condition of Pasquill F, the values of C_{HF} and Iu are smaller than the reference levels from 7600 m and 1500 m of the release point, what also classifies the event as an accident.

Scenario C2

According to criterion 1 for Class D of atmospheric stability, the event is classified as critical, because C_{HF} and Iu for individuals of the public (Area 3) exceed the values of GROUP 2 and at location (Area2) the values of C_{HF} and Iu surpass the values of GROUP 1. For class F of atmospheric stability, the event is classified as catastrophic, once C_{HF} and Iu for individuals of the public (Area 3) exceed the values of GROUP 1. Regarding the criterion 2, for the condition of atmospheric stability Pasquill D, near the limit of the property, C_{HF} is below 35 mg/m^3 and Iu is below 10 mg, likely to be classified as an abnormal event. In condition F, the results characterize accident, because C_{HF} is less than 35 mg/m^3 from 6700 m and Iu is below 10 mg from 1300 m to the point of release.

Scenario C3

In relation to criterion 1, it is observed that the toxicological consequences due to exposure to C_{HF} and Iu are classified as: marginal for Class D of atmospheric stability, as C_{HF} and Iu on the leakage area (Area 1) are higher than the value of GROUP 1; critical for the class F of atmospheric stability, because C_{HF} and Iu on the locality (Area 2) is higher than in GROUP 1. Considering the classification criterion 2, the results indicate abnormal event for the Class D of atmospheric stability, as C_{HF} is less than 35 mg/m^3 from 350 m and Iu less than 10 mg from 500 m. In class F of atmospheric stability, there are attained levels of more than C_{HF} 35 mg/m^3 at distances up to 1600 m, characterizing accident. Iu is less than 10 mg to 960 m from the point of release.

4. Conclusions

The broadest criterion for consequences evaluation on human health due to the event of accidental release of UF_6, as: despicable, marginal, critical or catastrophic, is the widest because it involves the evaluation of consequences in three different areas: the area of occurrence of leakage (Area 1), the location (Area 2) and the area outside the limits of the property (Area 3). The toxicological consequences for scenario C1, Class D of atmospheric stability and scenarios C1 and C2, considering the class F of atmospheric stability, are classified as catastrophic, once they represent situations where individuals of the public are subject to C_{HF} or Iu exceeding the values in GROUP 1. It should be noted, however, that the scenario C1 is extremely conservative, as it does not take into consideration the performance of the ventilation system filters, disregards the confinement exerted by the building itself and does not include the phenomena of elevation of the plume and deposition of particles of UO_2F_2 on the floor of the building. If these factors are added with the low rates of failure of UF_6 pipes and filters, plus the fact that operators of the facility could serve to isolate the area affected by the leakage, the C1 scenario can be considered not credible. It follows, therefore, that the risk of C1 associated with individuals of the public is almost negligible.

The second criterion used to classify events of accidental release of UF_6, only classifies the toxicological consequences on the health of individuals from the public. It has only two categories in the classification of the consequences: abnormal event or accident. It is verified therefore, that the worst conditions of release correspond to the F condition of atmospheric stability, since for all examined cases the consequences are classified as accidents. The scenario C3 illustrates how the adoption of impact attenuators measures (such as the performance of the ventilation system filters) may benefit the plant and reduce the risks to individuals from the public. From the obtained results, it can be observed that the criteria 1 and 2 are coherent, as there are equivalences between the catastrophic or critical categories on criterion 1 and the accident classification on criterion 2.

5. Acknowledgements

The authors acknowledge and thank the institutional support offered by Centro Universitário da FEI.

6. References

Hannaman, G. W., Kryskas, P., Mahan, J., 1995. Qualitative Methods for Assessing Risk. ASME, USA.

Molnary, L.,1993. Caracterização de um modelo de camada limite planetária para avaliar liberações de radionuclídeos em instalações nucleares. Dissertação (Mestrado)–IPEN. São Paulo: Universidade de São Paulo.

Nuclear Regulatory Commission, 1975. Sequoyah Uranium Hexafluoride Plant: Final Environmental Statement; NUREG 75 007.

Nuclear Regulatory Commission, 1986. Assessment of the Public Health Impact from the Accidental Release of UF_6 at the Sequoyah Hexafluoride Fuel Corporation Facility at Gore; NUREG 1189, v. 1.

Nuclear Regulatory Commission, 1991. Chemical Toxicity of Uranium Hexafluoride Compared to Acute Effects of Radiation, NUREG 1391.

Siman-Tov, M. et al., 1984. Scenarios an Analytical Methods for UF_6 Releases at NRC Licensed Fuel Cycle Facilities. Oak Ridge, TN, USA, NUREG /CR 3139.

World Bank, 1985. Manual of Industrial Hazard Assessment Techniques. London.

1947

Modeling of Kinetics of Water Droplets Coalescence in Crude Oil Emulsion Subjected to an Electrical Field.

Antonio E. Bresciani[a], Candido F. X. de Mendonça[b], Rita M. B. Alves[a] and
Claudio A. O. Nascimento[a]

[a]LSCP/CESQ - Department of Chemical Engineering - Polytechnic School – University
of São Paulo, São Paulo, SP, Brazil.
[b] Escola de Artes, Ciências e Humanidades - University of São Paulo, São Paulo, SP,
Brazil.

Abstract

Water is used in petroleum desalting units to dilute and remove the salted water droplets
that the crude oil contains. The basic processes promote the coalescence of small
droplets of conducting water dispersed in a crude oil emulsion. In order to make
separation easier, the emulsion is distributed horizontally between two electrodes and
subjected to an electrical field, which generates an attractive force among the droplets,
promoting coalescence phenomena and further sedimentation. The main purpose of this
study is to reduce the demand of fresh water and the liquid effluent generation in
refineries. This paper presents a new model developed in order to calculate the droplets
velocity by using the balance of the acting forces. The model is able to determine the
droplets trajectory in order to define if they can be separated from the continuous phase.
Besides the deterministic approaches based on traditional equations, the model uses also
the concept of cellular automata. Thus it is possible to solve the problem in a stochastic
way and to show visually the sequence of droplets collisions and coalescence
phenomena. This methodology enables to calculate the amount of water that can be
separated of the emulsion for a number of different operating conditions and then to
optimize the process. Comparisons between the obtained results by the developed
model and the operational performance of a real desalting unit are carried out. A good
accuracy is observed, which shows that the real process is very well represented by the
developed model.

Keywords: water use optimization, emulsion, coalescence, automata cell

1. Introduction

Studies in the optimization of water use in industrial process plants involve profound
knowledge of all processes that use water (Bagajewicz, 2000). It is necessary to
formulate equations to calculate water flow, contaminants concentrations at process
input and output and the maximum and minimum limits for these concentrations.
Desalting is an important process in petroleum refining that uses water to remove salt
that is present in petroleum when it is extracted in oil fields. Water is added to dilute
Sodium, Calcium and Magnesium Chlorides. An emulsion is formed, comprising
droplets with an average diameter in the range of 10 to 20μm (Bhardwaj, 1994). The
water separated from this emulsion removes salt in the petroleum, thus avoiding
corrosive wear on the equipment further downstream in the refining process, as well as

residual deposits and catalyst contamination. In order to facilitate the separation, an external electric field is set up which stimulates the agglutination of the smaller droplets into larger droplets that settle and are thus separated from the continuous phase. The industrial petroleum desalting system comprises of two desalting stages, involving the recycling of water, as demonstrated in Figure 1
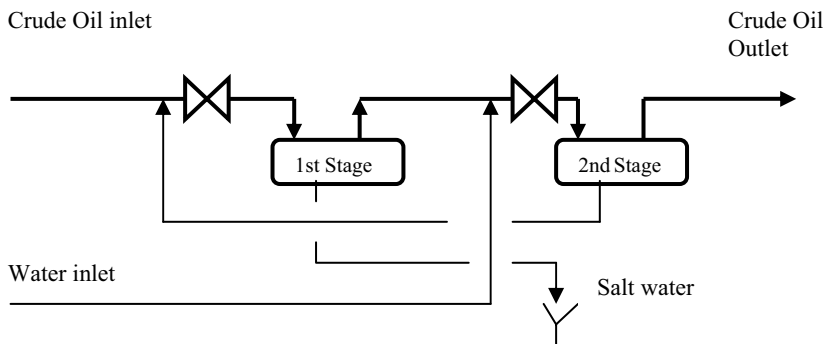
Crude Oil inlet   Crude Oil
                                                                                     Outlet

Figure 1. Desalting system employed by petroleum refineries.

These are normally modular, high-speed desalters, to which an external electric field is applied, and where the emulsion is injected into the space between the electrodes. Discharge from the continuous phase is considered to be horizontal in the volume between the electrodes and vertical after it passes by the electrodes, as demonstrated in Figure 2. The small droplets are affected by both the action of the electric field and gravitational force. They then agglutinate and are separated from the continuous phase. The smaller droplets that are not large enough to be decanted remain in the petroleum that exits the desalter.
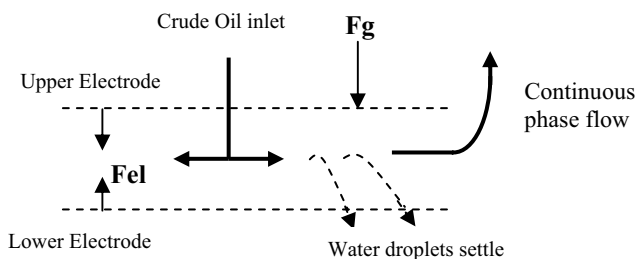

Figure 2. Discharge of the emulsion between the electrodes and forces that act upon the droplets.

The time in which the emulsion remains under the influence of the electric field is calculated by the equation for the speed of the horizontal discharge from the continuous phase, from the moment the emulsion exits the central distributor until it leaves the group of electrodes.
The balance of forces that act upon the droplets leads to the formulation of a mathematical model which enables the calculation of the time between collisions, the speeds immediately before collisions and the trajectory of the droplets.
The operational variables that have the most influence in this process are the operating temperature, the electric power gradient, the quantity of water in the mixture and the density and viscosity of the petroleum. The formulated model enables the study of the influence of these variables on the efficiency of the separation of the mixture, hence permitting process optimization.

The influence of the quality of water in the mixture is studied in laboratory experiments, where the separation time and the quantity of water separated are measured with analytical equipment that ascertains the intensity of light transmitted or refracted by the droplets.

2. Formulation of the mathematical model.

2.1. Horizontal discharge from the continuous phase
Equation (1) calculates the horizontal speed (v_h) of the continuous phase over the distance (z) from the distribution valve outlet to the desired position, for the flow and dimensions of the desalter studied.

$$v_h = 0.0278/(0.0235 + 0.322\, z + 1.10\, z^2)\ (m/s) \tag{1}$$

The time in which the emulsion is under the influence of the electric field is obtained through the integration of Eq. (1), and the result is 20 seconds.

2.2. Forces acting between droplets
The forces acting between droplets are the electric force (Fel), resulting from the effect of the electric field, gravitational force and viscous force. Natural attraction/repulsion forces (Van der Waals) and the Brownian force are not considered, as their intensities are much smaller than the others under consideration (Eow, 2001). Figure 4 demonstrates how the forces act upon the droplets.

F_{an} : Natural attraction inter-droplets force (van der Walls and repulsion)

F_{el} : Attraction inter-droplets force by electrical field

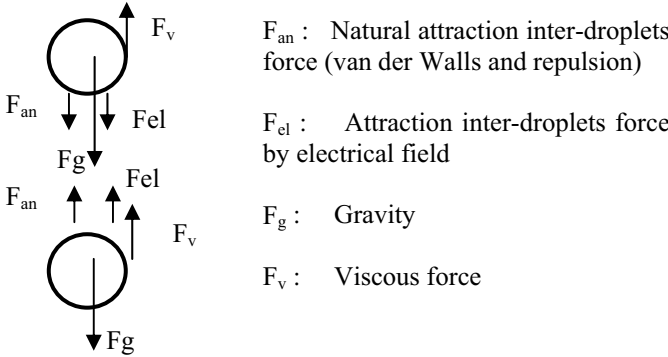F_g : Gravity

F_v : Viscous force

Figure 3 Forces acting on the droplets

The balance of forces acting upon each droplet is established, resulting in equation. (2), for the upper droplet (subscript 1), and equation. (3), for the lower droplet (subscript 2).

$$F_g + F_{el} - F_v = m_1\, dv_1/dt = 4/3\ \pi\, a_1^3\, \rho_a\, dv_1/dt \tag{2}$$

$$F_g - F_{el} - F_v = m_2\, dv_2/dt = 4/3\ \pi\, a_2^3\, \rho_a\, dv_2/dt \tag{3}$$

Through the force equations and the parameter values for the oil studied (viscosity, density and dielectric constant), for T=120 ºC, which can be considered to be the operating temperature of the desalter, we get Eq. (4), for the upper droplet, and Eq. (5), for the lower droplet.

$$1.57 + 0{,}134\ 10^{-3}\ E_0^2\, a_1\, a_2^2\, \ell^4 - 7.3\ 10^{-6}\, a_1^{-2}\, v_1 = dv_1/dt \tag{4}$$

$$1.57 - 0{,}134\ 10^{-3}\ E_0^2\, a_1^2\, a_2\, \ell^4 - 7.3\ 10^{-6}\, a_2^{-2}\, v_2 = dv_2/dt \tag{5}$$

where E_0 is the applied electric power gradient (kV/cm), a_1 and a_2 are the radii of droplets 1 and 2, m and ℓ are the mass and the distance between the droplets.

Equations (5) and (6) can be solved numerically for the pair of droplets of radii a_1 and a_2 under the effect of the electric power gradient E_0. The integration time must be very small ($2\ 10^{-6}$ seconds), requiring a lot of time for computational processing. Considering that it is an almost stationary regime, with $dv_1/dt= dv_1/dt=$zero, for 120°C, we get equation. (6), for the upper droplet, and equation (7), for the lower droplet.

$$v_1 = 2.15\ 10^5\ a_1^{\ 2} + 1.83\ 10^1\ E_0^{\ 2}\ a_1^{\ 3}\ a_2^{\ 2}\ \ell^4 \tag{6}$$

$$v_2 = 2.15\ 10^5\ a_2^{\ 2} + 1.83\ 10^1\ E_0^{\ 2}\ a_1^{\ 2}\ a_2^{\ 3}\ \ell^4 \tag{7}$$

Equations (6) and (7) are also solved numerically, with a longer integration time (10^{-3} seconds). Table 1 shows the results of the solution of the systems of equations 4/5 and 6/7 for the temperature of 120°C. The errors resulting from the consideration of an almost stationary state are small, which justifies its use to obtain shorter computation processing times.

Table 1. Time between collisions t_c for each power gradient, 7% water, for 10μm drops.

	Eo = 0.35 kV/cm	Eo = 0.35 kV/cm	Eo = 0.35 kV/cm	Eo = 0.35 kV/cm	Eo = 0.35 kV/cm
t_c (s) (7/8)	1.16	0.57	0.142	0.063	0.035
t_c (s) (num sol 5/6)	1.20	0.59	0.147	0.065	0.037
Error (%)	3.3	3.4	3.4	3.1	5.4
Absolute error (s)	0.04	0.02	0.005	0.002	0.002

2.3. The mathematical model employing the concept of cellular automata.
A model based on the concept of cellular automata was developed, which enables the application of equations of the process being studied to the whole group of droplets, in a stochastic manner, in order to visualize the collision process and optimize the process variables.

The concept of cellular automata was introduced by von Neumann (1970) and consists of a sequence of many cells carrying discreet values arranged in a matrix of these cells and the alteration of these values depends on the values of the variables of the neighboring cells. Wolfram (1983), in his discussion of cellular automata as systems of statistical mechanics, established that physical systems containing many discreet elements with local interactions are conveniently modeled as cellular automata.

The application of cellular automata to more complex systems has only come into use more recently, due to the greater availability of computational resources in recent times. For example, we can cite Ohgai et al (2007), who applied this concept to study the effects of accidents (explosions) on a certain physical environment. The simulation space consists of a group of cells that comprise a geometric solid where the dimensions are parameters. In this case, the height is fixed and corresponds to the application space of the electric power gradient, and is divided into 1143 cells, each with 0.1 mm. Thus, a cell corresponds to a horizontal slice of the geometric solid, as shown in Figure 4. Initially, an emulsion of water in oil, with a distribution of droplet size and the concentration of water defined, is randomly injected into the first 254 cells, which are in front of the nozzle of the distribution valve. The emulsion is subjected to the attraction forces between the droplets for 20 seconds, which is the time for the emulsion to drain

horizontally and to exit the space between the electrodes and the effect of the power gradient. Eqs. (6) and (7) are used to calculate the speed of displacement of the droplets due to the great reduction in the computational processing time and the small loss of precision. Collisions lead to the occurrence of larger droplets, and these drain vertically downwards, and then decant and are separated from the mixture.
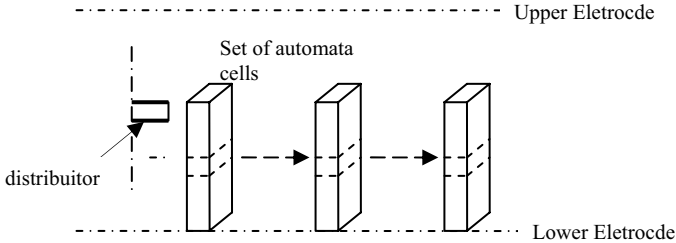

Figure 4. Diagram of the arrangement of the group of automaton cells in the space between the electrodes and their displacement in the Lagrangian perspective.

The state of any one cell depends on the location and the size of each droplet within the cell. The next state of a cell is calculated, and the size and location of each droplet of the cell is obtained. One droplet at a time is moved, and all the other droplets of the cell and the neighboring cells are "frozen". Each iteration consists of two steps. First, the displacement along axes x, y and z is determined for each droplet of a given cell. Vertical displacement on axis y is calculated from the speed resulting from the application of equations. (6) and (7) on all the surrounding droplets. The horizontal displacement results from the force introduced to represent the effects of alignment provoked by the electric power gradient and the turbulence resulting from the discharge of the mixture. Next, which droplets have suffered collision is ascertained, and the new size and position of the resulting droplet is determined.

In this study, each iteration corresponds to an interval of 1 millisecond, and twenty thousand iterations are necessary. Lastly, the volume of the small droplets remaining in the mixture is calculated, thus obtaining the efficiency of separation. Moreover, the group of cells is maintained geometrically fixed, in other words, horizontal and vertical displacement, due to the increase in drainage area resulting from the real geometrical shape of the desalting model, is not considered.

The model considers the variables of operational temperature, concentration of water and distribution of the size of the droplets as parameters, enabling the study of the influence of each one of these in the separation of the emulsion. Due to low kinetic energy the droplets with a radius of less than 5μ do not agglutinate when they hit others that are also less than 5μ. For the case under study, it is also considered that the droplets formed with a radius more than 88 μm may be separated from the mixture by gravity.

3. Results

Figure 5 shows that the efficiency of dehydration, which is the calculation of the percentage of separated water, compared with the volume of water in the emulsion at entry to the desalter, increases with the increase in the concentration of water. The BSW (Basic Sediment and Water), that is the content of water in the mixture that leaves the desalter, reduces with the increase in the concentration of water. Both results are in accordance with expectations for the process because, with the increase in the concentration of water, the number of droplets increases and the distance between the droplets is reduced, resulting in the collision between the droplets becoming more

intense. Figure 6 shows that the increase in the intensity of the power gradient applied has as a consequence the reduction in the BSW and increase in dehydration efficiency.
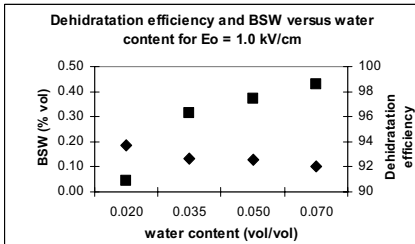

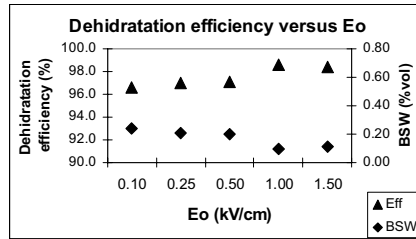Figure 5. Dehydration Efficiency due to water concentration

Figure 6. Dehydration efficiency due to electric voltage gradient.

Figure 7 shows emulsion situations, recorded from a computer screen and application stages of the model, thus enabling the determination of the evolution of the agglutination process and the separation of droplets.
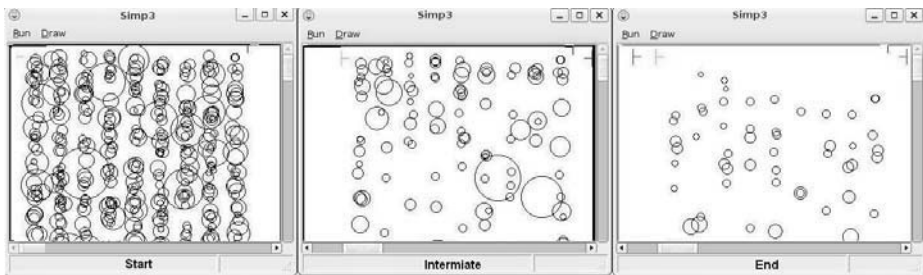
Figure 7. Recorded from process computer screens showing the application of the model.

4. Conclusion

The model developed based on cellular automata with balance of force equations enables the simulated accompaniment of the physical process involving the sequence of agglutination. The results obtained with application of the model based on cellular automata show that the influence of each operational variable studied is coherent with the fundamental theory of the process and they are similar to the results of tests executed in an industrial plant.

References

Bagajewicz, M. A review of recent design procedures for water networks in refineries and process plants. Computer and Chemical Engineering. 24, 2093 – 2113, 2000.

Bhardwaj, A. and Hartland, S. Dynamics of emulsification and demulsification of water in crude oil emulsions, Ind. Eng. Chem. Res, 33, 1271-1279, 1994.

Eow, J.S., Ghadiri, M., Sharif, A.O., Williams, T.J. Electrostatic enhancement of Coalescence of water droplets in oil: A Review of the current understanding, Chemical Engineering Journal, 84 173-192, 2001.

J. von Neumann, The General and Logical Theory of Automata, in J. von Neumann. Collected Works (ed. A.H. Taub), Vol. 5, p.288; J. von Neumann, ``Theory of Self-Reproducing Automata'', (ed. A.W. Burks), Univ. of Illinois Press (1966); ed. A.W. Burks, ``Essays on Cellular Automata'', Univ. of Illinois Press (1970).

Ohgai, A., Gohnai, Y., Watanabe, K., Cellular automata modeling of fire spread in built-up areas- A tool to aid community-based planning for disaster mitigation, Computers, Environment and Urban Systems, 31 441-460, 2007

S. Wolfram, Statistical Mechanics of Cellular Automata, Rev. Mod. Phys. 55 (1983) 642.

On the Optimal On-Line Management of Photovoltaic-Hydrogen Hybrid Energy Systems

Victor M. Zavala,[a,b] Mihai Anitescu,[a] and Theodore Krause[b]

[a]*Mathematics and Computer Science Division, Argonne National Laboratory,*
  9700 S. Cass Ave, Argonne, IL, USA
[b]*Chemical Technology Division, Argonne National Laboratory,*
  9700 S. Cass Ave, Argonne, IL, USA

Abstract

We present an on-line management strategy for photovoltaic-hydrogen (PV-H_2) hybrid energy systems. The strategy follows a receding-horizon principle and exploits solar radiation forecasts and statistics generated through a Gaussian process model. We demonstrate that incorporating forecast information can dramatically improve the reliability and economic performance of these promising energy production devices.

Keywords: receding horizon, stochastic, Gaussian process, solar, hydrogen.

1. Introduction

Hybrid technologies are attractive alternatives for satisfying increasing energy needs in diverse industrial sectors. The main idea is to couple components that generate power from different sources such as fossil fuels or renewables. With this approach, it is possible to overcome cost and efficiency limitations of the individual components and, in turn, minimize the overall system costs. A promising hybrid is the so-called photovoltaic-hydrogen (PV-H_2) system. A schematic representation is given in Figure 1. The idea is to generate electricity from solar radiation to fulfill a given load. The excess power is stored in a battery bank or in the form of hydrogen produced by electrolysis. The stored hydrogen can be converted back to electric power by using a fuel cell system. Power conditioning devices are used to regulate the voltages of the different devices, which are connected to a common busbar.

An important obstacle affecting the reliability of PV-H_2 systems is the fact that the main energy source is intermittent and highly uncertain. To illustrate this, in Figure 2 we present the total solar radiation for year 2004 at location 41° 59' N/87° 54' W in the Chicago, IL area. Another important issue is the fact that the components might have significantly different efficiencies (giving rise to different levels of power losses). Consequently, it might not be immediately evident which component is the optimal one to store and provide energy at a particular time. Motivated by these issues, researchers have devoted significant effort to developing on-line control or management strategies. Most of the strategies reported so far have been based on fuzzy logic and neural networks techniques (Vosen and Keller 1999, Ulleberg 2004). While these strategies might seem practical at a first glance, they are not general enough to handle economics, forecasts, and operational limits systematically. In this work, we present a general on-line management strategy for PV-H_2 hybrid systems. The strategy follows a receding-horizon (RH) technique and incorporates an economic objective function. With this strategy, we can directly study the effect of using forecast information on the overall operating costs. We demonstrate that using management strategies that neglect the

future radiation trends can severely affect the system costs and reliability. Motivated by this observation, we derive a strategy to obtain approximate forecasts and associated statistics through a Gaussian process modeling technique. We then use this information to derive a stochastic RH strategy that satisfies the load reliably.
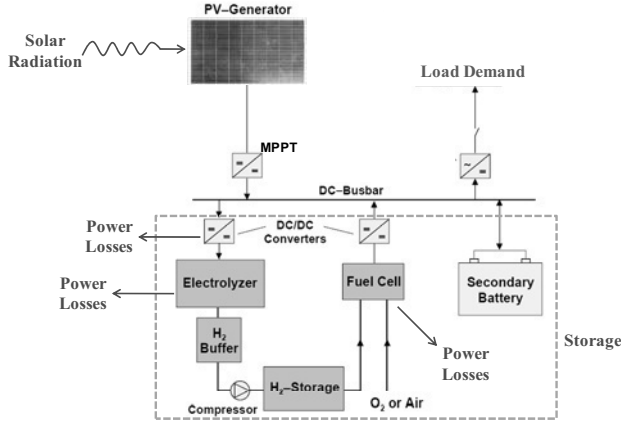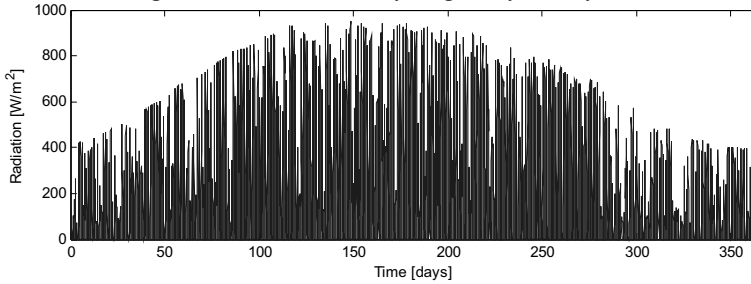

Figure 1. Photovoltaic-Hydrogen Hybrid System

Figure 2. Profile of total solar radiation in Chicago IL, 2004.

2. System Dynamic Model

The dynamic model comprises a systems wide power balance. The power entering through the solar module at time t_k is denoted by P_k^{PV}(kW). This can be calculated by using the measured radiation G_k^T (kW/m^2) and the module design characteristics. The electric current will go through a DC-DC converter that will seek to match the electric current voltage to the voltage of the distribution busbar. This conditioning process has an inherent efficiency θ_{PV} and generates power losses. The remaining power $\theta_{PV} P_k^{PV}$ is sent to the busbar to satisfy the current load P_k^{LOAD}. The excess power can be used to produce hydrogen in the electrolyzer and/or to charge the battery. In order to run the electrolyzer, the power extracted P_k^{EL} passes through a buck DC-DC converter, which brings the current voltage down to the operating voltage of the electrolyzer. The efficiency of this step is θ_{BU}. The remaining power $\theta_{BU} P_k^{EL}$ enters the electrolyzer. The conversion process to hydrogen has an efficiency θ_{EL}. The *net price* for each power unit produced by the electrolyzer is given by C_{EL}. Since hydrogen can be seen as an asset, the net price can be negative. The produced power $\theta_{BU}\theta_{EL} P_k^{EL}$ in the form of hydrogen is stored in a storage system modeled by a difference equation of the form $E_{k+1}^{H2} = E_k^{H2} + \Delta_k (\theta_{BU}\theta_{EL} P_k^{EL} - P_k^{FC})$, where E_k^{H2} is the total energy stored (kWh) at time t_k and $\Delta_k = t_{k+1} - t_k$ (hr). The hydrogen *state of charge* is defined as $SOC_k^{H2} =$

$100 \frac{E_k^{H2}}{E_{MAX}^{H2}}$, where E_{MAX}^{H2} is the nominal maximum capacity (kWh). A certain amount of power P_k^{FC} can be withdrawn from the storage to feed a fuel cell and generate electric power. The conversion process has an efficiency θ_{FC}. The cost for each unit of power produced by the fuel cell is given by C_{FC}. The remaining power is then passed through a boost DC-DC converter that brings the voltage of the current up to the operating voltage of the busbar. The process has an efficiency θ_{BO}. The remaining power $\theta_{FC}\theta_{BO}P_k^{FC}$ is sent to the distribution busbar. The system might be able to buy a given amount of power P_k^G from the grid in order to balance the system. This power will have a cost C_G which depends on the location and the degree of independence required by the application (e.g., for a stand-alone system, $C_G = \infty$). Excess power at the busbar can also be dumped to the grid or environment, which is modeled by variable P_k^D. The cost of dumped power is C_D. If the power is dumped to the grid, this cost becomes an asset (set by net-metering rates). The power remaining at the busbar can be used to either charge or discharge the battery. The net battery power P_k^B is calculated as $P_k^B = \theta_{PV}P_k^{PV} + P_k^G + \theta_{FC}\theta_{BU}P_k^{FC} - P_k^{EL} - P_k^{LOAD} - P_k^D$. The battery balance is $E_{k+1}^B = E_k^B + \Delta_k P_k^B$, and the state-of-charge is $SOC_k^B = 100 \frac{E_k^B}{E_{MAX}^B}$. The fixed model inputs are P_k^{PV} and P_k^{LOAD}. The degrees of freedom are P_k^{EL}, P_k^{FC}, P_k^G, and P_k^D.

3. Management Strategy

The RH strategy solves, at time t_k, a linear programming (LP) problem of the form

$$\min_{P_j^{EL}, P_j^{FC}, P_j^G, P_j^D} \sum_{j=k}^{k+N-1} C_{EL}P_j^{EL} + C_{FC}P_j^{FC} + C_G P_j^G + C_D P_j^D$$

$$s.t. \; E_{j+1}^{H2} = E_j^{H2} + \Delta_j \left(\theta_{BU}\theta_{EL}P_j^{EL} - P_j^{FC} \right), \quad j = k, \dots, k+N-1$$
$$E_{j+1}^B = E_j^B + \Delta_j P_j^B, \quad\quad\quad\quad\quad\quad j = k, \dots, k+N-1$$
$$E_k^{H2} = given, \quad E_k^B = given$$
$$P_j^B = \theta_{PV}P_j^{PV} + P_j^G + \theta_{FC}\theta_{BU}P_j^{FC} - P_j^{EL} - P_j^{LOAD} - P_j^D, \; j = k, \dots, k+N-1$$
$$0 \geq SOC_j^B \leq 100, \quad 0 \geq SOC_j^{H2} \leq 100, \quad j = k, \dots, k+N.$$

From the solution of the LP, we obtain the optimal future trajectory for the electrolyzer, fuel cell, grid, and dump powers that minimizes the operating costs, maximizes H₂ production and, simultaneously, satisfies the load and the storage limiting levels. To solve the LP, we need information of the future solar power $P_j^{PV}, j = k, \dots, k+N-1$ expected to be available. Important research questions that, to the best of our knowledge, have not been addressed so far are: What is the economic impact of folding forecast information in on-line management strategies? What is an appropriate forecast horizon? How can we get accurate forecast information? To address these questions, we perform a numerical case study. The efficiencies of the components are obtained from Vosen and Keller (1999). The unit costs are obtained from Stoll and von Linde (2000). A constant load of 1kW is assumed. The maximum peak PV power is 5 kW. We first solve an open-loop optimal control problem using *perfect* forecast information for a horizon of 365 days ($N = 365x24 = 8760$ hours). We use the optimal cost as a reference for the *best economic performance possible* over one year of operation. This is on the order of \$1,000/yr. We then run the closed-loop RH strategy spanning the year for different horizons $N = 1, 6, 12, 24, 3x24, 7x24, 14x24$ (hr) with an update time of $\Delta_k = 1$ hr and compute the corresponding relative costs. This required extensive computations. For each scenario, approximately 8,500 LP problems needed to be solved. The 14-day forecast LP contains 2,000 constraints and 1,000 degrees of freedom

and can be solved in less than one second with a state-of-the-art solver. The results are summarized in Figure 3. Several interesting and unexpected conclusions can be drawn from this study: (1) the relative operating costs decay exponentially to zero as the horizon is increased; (2) for a purely reactive strategy ($N = 1$ hr), the relative costs can go as high as 300%; and (3) the *overall best cost can be obtained with a finite forecast* ($N = 24x14$ hr). This last result has important practical implications because it is often difficult to obtain accurate long-term weather forecasts. In addition, note that the economic penalty of using a short forecast of 24 hr is just an increase of 10% in relative costs, whereas the penalty for a forecast of 12 hr is 31%. This implies that a practical horizon should be sufficiently long to capture the periodicity of the daily radiation. The reason for these strong economic penalties becomes evident from Figure 4. Here, we present the power profiles for the fuel cell for both the open-loop and the $N = 12$ hr closed-loop cases. Note that *shorter forecasts induce more aggressive control actions,* which in turn affect the costs. As we increase the horizon, the system is allowed to react more *proactively,* which is reflected in smoother controls.
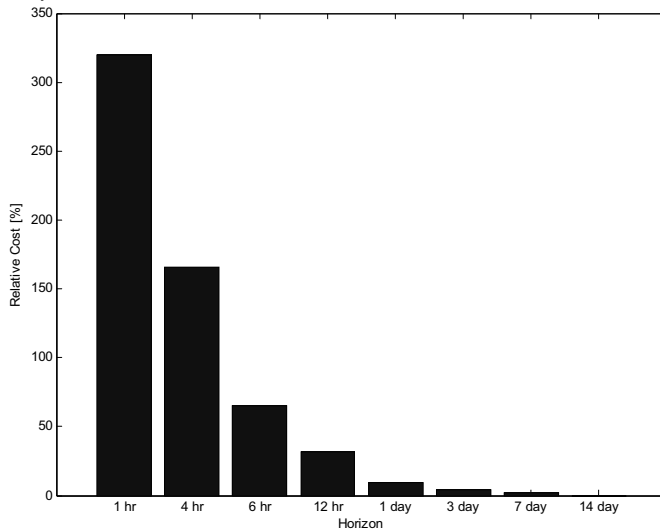

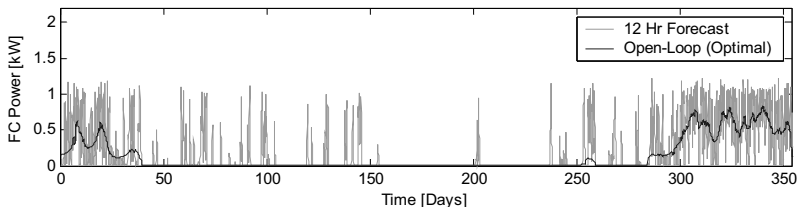Figure 3. Impact of forecast horizon on economic performance.

Figure 4. Impact of forecast horizon on power profiles.

4. Gaussian Process Model and Stochastic Management Strategy

Gaining access to solar radiation predictions can be complicated or impractical. In addition, if the forecast is not accurate enough, the management strategy can run out of stored energy prematurely and will be unable to satisfy the load. This situation is particularly critical in stand-alone systems. In the absence of forecasts, we could assume that the radiation profile of the next day will be similar to that of the previous day. Another option is to use historical data to construct a dynamic empirical model. For instance, a time-series approach could be used to build an auto-regressive (AR) model.

An approach that has recently received attention is Gaussian process (GP) modeling (Rasmussen and Williams 2006). The idea is to construct an AR model by specifying the structure of the covariance matrix rather than the structure of the dynamic model itself. We have found that this feature makes the GP modeling particularly flexible. Consequently, this was the approach used in this work. Because of space restrictions, we present only the basics of the GP algorithm. We construct a model by regressing the future radiation (output) y_{k+1} to the current radiation and to the radiation observed T time steps ago (we use $T = 24$ hr to enforce periodicity). We define the inputs $x_k = [y_k, y_{k+1-T}]$ to give $y_{k+1} = f(x_k)$. We collect a number of input-output pairs as training data sets represented by Y and X. We assume that the inputs are correlated through an exponential covariance function

$$K(x, x', \vartheta) = \vartheta_1 \exp\left(-\frac{1}{\vartheta_2}\|x - x'\|^2\right) + \vartheta_3,$$

where $\vartheta_1, \vartheta_2, \vartheta_3$ are parameters estimated by maximizing the log likelihood function:

$$\log p(Y|\vartheta) = -\frac{1}{2} Y K^{-1}(X, X, \vartheta) Y - \frac{1}{2}\log \det(K(X, X, \vartheta)).$$

Once the parameters are obtained, we can compute mean predictions Y^T with associated covariance K^T at a set of test points X^T. In our context, these are the time-varying radiation trends. The predictive equations are

$$Y^T = K(X^T, X, \vartheta) K^{-1}(X, X, \vartheta) Y$$
$$K^T = K(X^T, X^T, \vartheta) - K(X^T, X, \vartheta) K^{-1}(X, X, \vartheta) K(X, X^T, \vartheta).$$

In Figure 5, we present the mean forecast and 100 samples drawn from the normal distribution $N(Y^T, K^T)$ at a particular day. We use approximately 400 training data sets. Note that the distribution captures the true radiation values, implying that the assumed covariance structure is reasonable. We use the GP forecast distributions to derive a stochastic RH strategy that minimizes the *expected* cost and satisfies the constraints for all possible realizations of the future radiation. The objective function takes the form:

$$\min_{P_j^{EL}, P_j^{FC}, P_j^G, P_j^D} \quad E\left[\sum_{j=k}^{k+N-1} C_{EL} P_j^{EL} + C_{FC} P_j^{FC} + C_G P_j^G + C_D P_j^D\right].$$

Symbol $E[.]$ denotes the expectation operator. We solve this infinite-dimensional problem using a sample-average approximation strategy. To test the stochastic strategy, we construct scenarios by sampling the predicted GP distribution. In Figure 6, we compare the performance of the optimal open-loop strategy with perfect forecast (RH optimal) with a strategy that uses the mean forecast (GP mean) and one that uses 100 samples (GP samples). The GP samples LP contains 16,000 constrains and 5,000 degrees of freedom and can be solved in 3 seconds. In addition, we analyze the performance of a simple strategy that propagates the radiation of the last day to the next day (RH simple). In the top graph, we present the state-of-charge of H_2 storage along the year. Note that the GP mean and GP samples approach obtain similar levels to those obtained by RH Optimal. On the other hand, the total H_2 produced with RH Simple is 10% lower. In the bottom graph, note that GP mean fails to satisfy the power demand at one instance during the year (indicated as large or infinite cost). RH simple is the least reliable strategy; failing at 4 instances (overlapping due to time scale). On the other hand, the stochastic GP samples strategy is always able to satisfy the demand.

5. Conclusions

In this work, we have derived a receding-horizon strategy to perform the on-line management of PV-H_2 hybrid energy systems. We conclude that a few days long forecast is sufficient to obtain an acceptable economic performance. In addition, we demonstrate that capturing the uncertainty of the future radiation trends is critical to satisfy the load reliably.
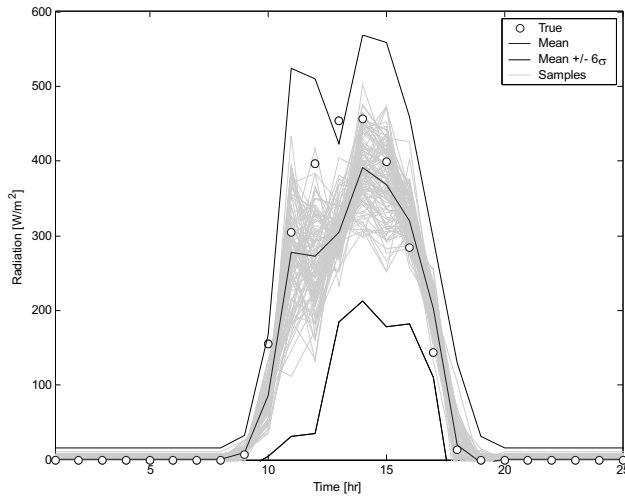
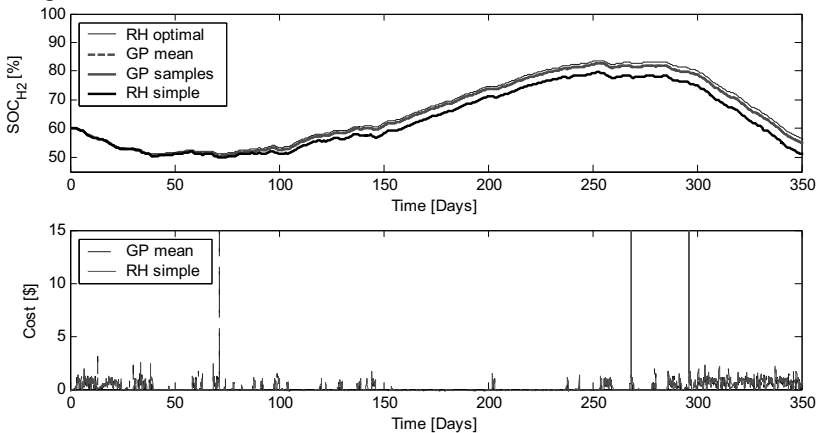Figure 5. Mean forecast and 100 realizations obtained with the GP model.

Figure 6. Performance of stochastic RH strategy with GP forecasts.

Acknowledgment

References

O.Ulleberg, 2004. The importance of control strategies in PV–hydrogen systems, Solar Energy, 76, 323-329.

S. R. Vosen and J. O. Keller, 1999. Hybrid energy storage systems for stand-alone electric power systems: optimization of system performance and cost through control strategies, Int. J. Hydrogen Storage, 24, 1139-1156.

R. Stoll and F. von Linde, 2000. Hydrogen - what are the costs? Hydrocarbon Processing, 79, 42-46.

C.E. Rasmussen and C.K. Williams, Gaussian Processes for Machine Learning. 2006, MIT Press.

Operator Trainer System for the Petrobras P-26 Semi-Submersible Oil and Gas Production Unit

A. C. Pereira,a A. Riera,b , G. Padilla,b E. Musulin,b N. J. Nakamurac

aPetrobras, Av. Chile 65, Rio de Janeiro 20031-912, Brasil
bSoteica Lationamerica S.A., A. Thomas 796 6B, Buenos Aires C1427CCU, Argentina
cSOteica do Brasil Ltda., Rua Dr. Cid de Castro 108, São Paulo 04064-040, Brasil

Abstract

Operator trainer systems aim to improve operator performance, by simulating scenarios such as emergency conditions, thus reducing accidents and increasing processes economical results. In this paper, we present PETROBRAS' Oil & Gas Production Process and Utilities Simulator Environment called AMBTREI (Training Environment) that mimics the actual Control Room of an E&P semi-submersible Platform at a very high fidelity level. This training environment was created utilizing Soteica's Operator Training System solution (S-OTS). The dynamic process model will be described as well as the Process Control Interface that was implemented. The software used will be explained in detail and the conclusions that have been reached in almost 2 years of use will be presented. **Keywords:** Operator Training, offshore platform, dynamic process simulation.

1. Introduction

1.1. Objectives

Operator trainer systems aim to improve operator performance thus reducing accidents and increasing processes economical results. They allow, among other capabilities:

- simulation of scenarios, such as emergency conditions.
- modeling for operator training purposes.
- studying plant behavior when an upset occurs

Petrobras' Project IND P&G 26.2.2 had the following main objectives:

- to increase capabilities at offshore E&P area, keeping PETROBRAS' Operational Excellence Plan (Plano de Excelência Operacional) guidelines according to its safety rules.

- to provide training, qualification and certification programs for professionals working at oil production areas (process and utilities) in offshore E&P platforms, focusing on normal and emergency operation techniques besides start-up and shutdown platform processes, avoiding on-site training risks.

1.2. Project Highlights

As part of the project, AMBTREI (Training Environment), an Oil & Gas Production Process Simulator for an E&P semi-submersible platform was created. Located at the Núcleo de Treinamento Offshore (NTO) Engenheiro Nelson Stavale Malheiro, in the Technology Center Euvaldo Lodi, Benfica, north part of Rio de Janeiro, this training environment aims to reproduce the actual Control Room of an E&P semi-submersible Platform at the highest fidelity level.

It allows training not only on oil processing but also utilities related processes. Being one of the most modern in the world, its implementation was possible through an agreement between PETROBRAS, Firjan/SENAI-RJ and SOTEICA Ideas and Technologies LLC, with representation in Brasil, acting as the Operator Training Software (S-OTS) provider and dynamic process model developer. Soteica brought to the project past experience in the field of Operator Trainers (Ref. 1) as well as a previous assignment with Petrobras' CENPES in the development of a dynamic model for the same platform (Ref. 2). PETROBRAS also took an active part in the project through its business units UN-BC, E&P-ENGP and CENPES, giving technical and consulting support.

AMBTREI main customer, PETROBRAS, will be training 1824 production operators for the first 7 years of the contract, achieving big economical savings. For Firjan/SENAI-RJ, this deal represents one more step forward its objective of becoming a training solutions provider for the Oil & Gas industry.

As part of the agreement between PETROBRAS and Firjan/SENAI-RJ, the latter expects to train platform operators from other Oil & Gas companies, so that, a partnership with ABERDEEN SKILLS AND ENTERPRISE TRAINING (ASET) was created, through which the Scottish company guarantees Firjan/SENAI-RJ exclusive rights to develop these training activities in Latin America.

1.3. Main Project characteristics

The following items describe the AMBTREI project:

• Production and Utilities Process dynamic model was based on P-26 PETROBRAS Platform. It's possible to simulate other platforms modifying this model.

•   This model takes into account:

   - Production and Utilities systems found on actual platforms such us: oil transportation to the platform through risers, three-phase oil-gas-water separation, oil exportation, gas compression, sea water lift, compressed air systems, hot and cold water systems among others.

   - Instrumentation as well as dynamic behavior of input and output variables

• Virtual Platform and Control System bi-directional interaction reproducing actual platform behavior.

• Virtual Control Room (supervisory system) and Control System bi-directional interaction, allows training the operators in using actual supervisory systems, mimicking actual platform behavior.

• Process and Utilities Production training courses are currently available with 2 options: Improvement and Certfication.

• SENAI-RJ instructors, trained and certified by ASET (Aberdeen, Scotland), are located in a trainers' room next to the Virtual Control Room. This room is equipped with Trainer Nodes allowing them to select preconfigured training scenarios, create new ones, simulate participation of field support teams to execute manual movements such as valves opening, etc.

• An ASET certification training course as well as SENAI-RJ certification is available for highly experienced Oil & Gas and Utilities Process Operators.

2. P-26 Platform Dynamic Process Model

The dynamic process model was developed using Hysys Dynamics, a first-principles based simulator. The process model includes:

• 16 production wells.   • Gas Booster compression.

- 10 injection wells.
- Production manifold gas exportation.
- Oil-water separation (electrostatic separation was also considered).
- Oil-gas separation.
- Re-injection Gas Compression.
- Glycol regeneration unit.
- Fuel gas system.
- Hot water system (2 furnaces).
- Sea water lift and distribution.
- Cooling water system.
- Water injection.

The first step in model development involved data gathering and other preparation issues. The following documentation was required:
- P&IDs
- Control system diagrams
- Operational procedures
- Input stream data (pressure, temperature and composition of well fluids arriving on P-26);
- Operating screen snapshots in normal processing condition;
- Main equipment data sheets (vessels, drums, pumps, compressors, valves, etc.);
- Pumps and compressors curves
- Piping data and equipments ground level when needed.
- Feed streams composition analysis data.

Turbine modeling data were especially difficult to obtain. The process data available were related to the global turbine performance only. Data for individual pieces (air compressor, combustion chamber, gas generator and power turbine) were not made available by the manufacturers as these data are considered confidential. A single set of intermediate information (in between pieces of equipment) was obtained. From this information set, additional data for turbine modeling were estimated. To get closer to reality, more intermediate data would be necessary.

Regarding other equipment, almost all required modeling data were gathered. Some minor equipment data sheets were missing and had their data estimated to reproduce actual plant behaviour.

3. Process Control Interface and Details

One of the main tasks of an OTS project is to define the interface between the dynamic process model and the control system. Different alternatives were taken into account:
- Emulation: simulation of the control system with similar look & feel. It eliminates system integration issues with the offline automation system and allows complete control of the simulation environment. However, emulated systems do not protect the integrity or fidelity of automation systems and do not support accurate implementation of the training system components
- Stimulation: usage of actual automation system components that accurately and exactly represent the control system behaviour.
- Virtual stimulation: control configuration and displays are identical to the actual DCS, but control logic runs on virtual controller hardware instead of actual controller hardware.
- Partial stimulation: actual automation system components may be used with some emulation of lower complexity parts of the automation system.

Partial stimulation was the adopted approach, replicating the original installation (hardware and configuration) whenever control hardware was available. Emulation was the choice at the following situations:

- minor equipments control programs run on same model and provider hardware, but to cut costs and reduce installation times, programs were migrated to run in main control hardware
- control hardware was difficult to get.
- some minor equipment logic was so simple that the cost-benefit ratio to add necessary hardware made the effort worthless.

4. Soteica Operator Trainer System (S-OTS)

The Soteica Operator Training System (S-OTS) aims to improve operator performance thus reducing accidents and increasing processes economical results. In particular, it is an excellent tool to train operators under process abnormal conditions.
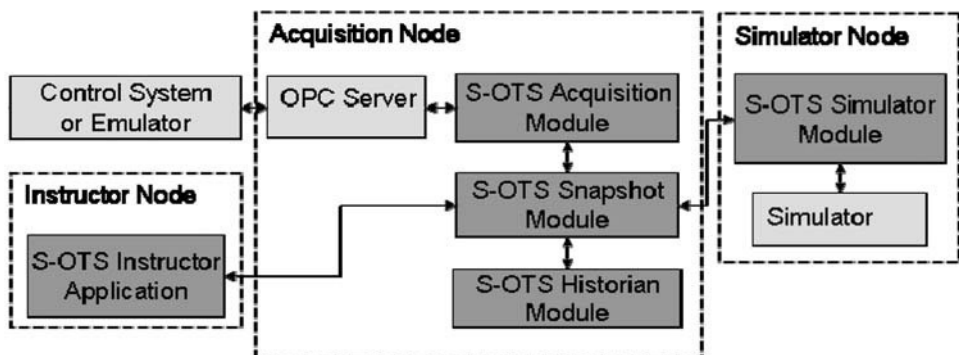
Figure 1: Soteica Operator Trainer System architecture. Arrows shows the flow of information (i.e. the state of the variables) between modules.

The S-OTS is responsible for capturing the operator actions, through the control system or field operated devices (informed to the instructor), send this information to the simulator, execute an integration step of the simulator, take the state of the variables of the simulator and put this information back to the control system. The S-OTS will also save the information related to each training session and keep a database of each user and its training sessions' information. Such information can then be accessed by means of an Excel ® add-in in order to evaluate the operators' performance.
S-OTS was developed following a modular and distributed architecture (see Figure 1) that manages the interaction between every system component (control system, Human-Machine Interface (HMI), Process Simulator and Instructor environment).

S-OTS modules are driven by a special module, the "S-OTS Instructor Application", which has three components:
- the Instructor application administrator
- the Excel history lookup tool
- the Training sessions common database.

It has a graphical user interface that interacts with the instructor to register users as instructors or operators to be trained, keep a list of training sessions developed, start, stop and pause training sessions, enter process perturbations or changes, read process variables values and searches sessions by user or date criteria and access history files from Excel®.

The S-OTS Snapshot module keeps the state of all the configured variables required for the training session at any given time. This module allows the other modules to exchange the state of the variables of the training sessions.

The S-OTS Simulator module reads the controller outputs from the S-OTS Snapshot, puts them into the Hysys model, executes an integration step and puts the new simulated values in the S-OTS snapshot.

The S-OTS Acquisition module is a bi-directional interface from the S-OTS Snapshot to the PLC controller using OPC standard. It writes the PLC controller outputs to the snapshot module and writes the current simulated variables values back to the PLC.

The S-OTS Historian module stores process data (temperatures, flows, etc.), instructor variables and field operated devices values. Once the training session is finished the S-OTS Instructor application gets the archived data and stores it in a file on the Instructor Node. The access to historical data will be available through the Excel® history lookup tool.

As a consequence of its modular architecture, additional functionality could be added to the S-OTS system to satisfy special needs.

5. Conclusions

As result of this project, PETROBRAS has a Production Process and Utilities Simulator Environment called AMBTREI (Training Environment), that mimics the actual Control Room of an E&P semi-submersible Platform at a very high fidelity level, allowing among other benefits:

- to increase capabilities at offshore E&P area
- to provide training, qualification and certification programs, focusing on normal and emergency operation techniques besides start-up and shutdown platform processes, avoiding on-site training risks and achieving important economical savings.
- to improve operators performance reducing accidents and increasing process economical results. In particular, the S-OTS environment is specially suited to train operators under abnormal process conditions.
- Trained people in 2007 (activities started on Jul-2007):   90.
- Trained people until November 2008 (Nov-11-2008):   167.

It's important to emphasize the high level of acceptance that these training activities have, based on the Operators' feedback. In particular they pointed out the following aspects:

- The interactive way in which activities are carried, and
- as this activity covers a lot of very useful topics it serves as a refresher of a great amount of knowledge acquired in past training courses.

6. Acknowledgments

References

Rajeev Agarwal, Ky Alston, Thomas J. Oliver, Juan Pablo Ruiz, Sebastián Cúneo, Advanced Operator Training and Plant Operations for a Gas-Plant using First Principles Dynamic modeling and Human-Machine Interface Emulation, IAPG Natural Gas Treatment Congress, 2004,<http://www.soteica.com/files/papers2dwnld/2004_OTS_Gas_Plant_IAPG_Natural_Gas _Treatment_Congress.pdf>

E. N. Satuf, J. C. Vaz1, R. F. Pimenta, R. Agarwal, How to operate a complex floating production unit inside a meeting room, 4th Mercosur Congress on Process Systems Engineering, 2005, <http://dpi.eq.ufrj.br/Anais_A/CENPROMER2005/nukleo/pdfs/1061_mpv_p26_enpromer_10 61_2005_final.pdf>

1965

Stochastic Programming with Tractable Mean–Risk Objectives for Refinery Planning Under Uncertainty

Cheng Seong Khor, Thi Huynh Nga Nguyen

Chemical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

Abstract

The application of information technology (IT) and information systems (IS) have been crucial in enhancing the operating flexibility and resiliency of refineries. In particular, the process systems engineering (PSE) community has been instrumental in carrying out a key role in extending the systems engineering boundaries from mere chemical process systems to the incorporation of business process systems with consideration for risk. Thus, this paper considers a robust framework for the economic and operational risk management of a refinery under uncertainty by extending an existing two-stage stochastic program with fixed recourse via scenario analysis. The problem is mathematically formulated as a two-stage stochastic nonlinear program with a tractable mean–risk structure in the objective function. Two measures of risk are considered, namely the metrics of mean-absolute deviation (MAD) and Conditional Value-at-Risk (CVaR). The scenario analysis approach is adopted to represent uncertainties in three types of stochastic parameters, namely prices of crude oil and commercial products, market demands, and production yields. However, a large number of scenarios are required to capture the stochasticity of the problem. Therefore, to circumvent the problem of the resulting large-scale model, we implement a Monte Carlo simulation approach based on the sample average approximation (SAA) technique to generate the scenarios. A statistical-based scenario reduction strategy is applied to determine the minimum number of scenarios required yet still able to compute the true optimal solution for a desired level of accuracy within the specified confidence intervals. The proposed model is illustrated through a representative numerical example, with computational results demonstrating how risk-averse-and risk-inclined solutions in the face of uncertainty can be attained in a risk-conscious model.

Keywords: two-stage stochastic programming, refinery planning, mean-absolute deviation (MAD), Conditional Value-at-Risk (CVaR)

1. Introduction

Stochastic programming has emerged as one of the most prominent operation research models for optimization involving uncertainty. Refinery planning problems are subject to uncertainty in many factors, which primarily includes fluctuations in prices of crude oil and saleable products, market demand for products; and production yields. The risk terms in Khor (2007) are handled using the metric mean-absolute deviation. After obtaining the first model with MAD as risk measurement, the second model is developed in which the risk terms are performed by CVaR. A comparison between the two models to assess which of these two risk measures is superior, both computationally

and conceptually, in capturing the economic and operating risk in the planning of a refinery. However, there are large numbers of scenarios that create difficulty to handle various circumstances. For example, there may be more than thousands of cases happening. It is hard to predict and control numerous scenarios. Therefore, it is necessary to find the minimum number of scenarios to capture all the circumstances. Monte Carlo simulation approach based on the sample average approximation (SAA) technique is applied in this thesis to generate the minimum number of scenarios which present for thousands cases.

Risk Model III as presented in Khor et al. (2007):

$$\max z = E[z_0] - \theta_1 V(z_0) - E_s - \theta_3 W \tag{1}$$

where $E[z_0]$ is the expectation of the original objective function z_0 with random price coefficients; θ_1 and θ_3 are weights representing the risk factors; $V(z_0)$ is the sampling variance of z_0; E_s is the expected recourse penalty, and W is the MAD-based risk measure. In this work, we extend model (1) by considering MAD as the risk measure in place of variance. Further, we investigate the viability of the risk measure Conditional Value-at-Risk (CVaR), which has gained wide attention in computational finance, within the domain of refinery planning. All the uncertain parameters are assumed to be discrete random variables.

2. Monte Carlo Simulation Approach based on Sample Average Approximation (SAA) Method

In this work, we adopt the Monte Carlo simulation approach for scenario generation based on the Sample Average Approximation (SAA) method (Shapiro, 2000; Shapiro and Homem-de-Mello , 1998; You and Grossmann, 2008). The procedure involved is as follows:

Step 1. A relatively small number of scenarios (for example, 50 scenarios) with their associated probabilities are randomly and independently generated for the uncertain parameters of prices, demands, and yields. This is accomplished by employing the Monte Carlo simulation approach based on the SAA technique. (This data is otherwise obtained from plant historical data.) The resulting stochastic model (a linear program) with the objective function given in (2) is solved to determine the optimal stochastic profit with its corresponding material flowrates.

$$\max E_z = E_{z_0} - E_\xi \tag{2}$$

where $\quad E_{z_0} = \sum_{i \in I} \sum_{s \in S} p_s c_{i,s} x_i$

$$E_\xi = \sum_{i \in I} \sum_{s \in S} p_s \xi_{i,s} = \sum_{i \in I} \sum_{s \in S} p_s \left[\left(c_i^+ z_{i,s}^+ + c_i^- z_{i,s}^- \right) + \sum_{k \in K} \left(q_{i,j}^+ y_{i,k,s}^+ + q_{i,j}^- y_{i,k,s}^- \right) \right]$$

Step 2. The Monte Carlo sampling variance estimator is determined using the optimal stochastic profit and flowrates computed in step 1.

$$S(n) = \sqrt{\frac{\sum_{s=1}^{S} \left(E_z - z_{i,s} \right)^2}{S - 1}} \qquad \text{where } z_{i,s} = \sum_{i \in I} \left(c_{i,s} x_i + \xi_{i,s} \right) \tag{3}$$

Step 3. The lower- and upper-confidence limits of the 95% confidence interval H of $1-\alpha$ are computed as follows:

$$\left[E_z - \frac{z_{\alpha/2}S(n)}{\sqrt{S}}, E_z + \frac{z_{\alpha/2}S(n)}{\sqrt{S}} \right] \tag{4}$$

Step 4. The minimum number of scenarios N that is theoretically required to obtain an optimal solution is determined using the relation below:

$$N = \left[\frac{z_{\alpha/2}S(n)}{H} \right]^2 \tag{5}$$

where the standard normal random variable $z_{\alpha/2} = 1.96$ at confidence interval $(1-\alpha) = 95\%$.

3. Formulation of Stochastic Refinery Planning Model with Mean-Absolute Deviation (MAD) as Risk Measure

The risk metric mean-absolute deviation (MAD) is employed as a measure of deviation from the expected profit (Konno and Koshizuka, 2005; Konno and Yamazaki, 1991). It is defined as follows:

$$\text{MAD}(x) = E \left| \sum_{j=1}^{n} R_j x_j - E \sum_{j=1}^{n} R_j x_j \right| \tag{6}$$

In this work, the rate of return R in (6) refers to unit cost of materials (crude oil and refinery products) and the amount of money x_j invested in an asset j refers to the refinery production amount. Therefore, the formulation of the MAD-based risk measure for price uncertainty becomes:

$$\text{MAD}(z_0) = \sum_{s \in S} p_s \left| \sum_{i \in I} c_{i,s} x_{i,s} - \sum_{i \in I} \sum_{s \in S} p_s c_{i,s} x_{i,s} \right| \tag{7}$$

while for demands and yields uncertainty, it is given by:

$$\text{MAD}_\xi = \sum_{s \in S} p_s \left| \sum_{i \in I} \xi_{i,s} - \sum_{i \in I} \sum_{s \in S} p_s \xi_{i,s} \right| \tag{8}$$

4. Formulation of Stochastic Refinery Planning Model with Conditional Value-at-Risk (CVaR) as Risk Measure

CVaR, also termed as Mean Excess Loss, Mean Shortfall or Tail VaR, is a risk assessment technique that is originally intended to be employed for reducing the probability that an investment portfolio will incur high losses. It offers the advantage of a linear programming formulation for determining the optimal solution of financial planning problems that explicitly minimizes loss or risk. CVaR is performed by taking the likelihood (at a specific confidence level, e.g., 0.95 or 0.99) that a specific loss will

exceed the metric known as Value-at-Risk (VaR). From a mathematical point of view, CVaR is derived by taking a weighted average between VaR and the losses exceeding VaR. For a discrete probability distribution function, CVaR can be defined as follows (Rockafellar and Uryasev, 2000; 2002):

$$F_\alpha(x, \text{VaR}) = \text{VaR} + \frac{1}{1-\alpha} \sum_{i \in I} \sum_{s \in S} p_s \left(f(x, y_{i,s}) - \text{VaR} \right) \tag{8}$$

Using CVaR as the risk metric yields the following form of the objective function:

$$\max z = E[z_0] - \theta_1 \text{CVaR}_{z_0} E_{s'} - \theta_3 \text{CVaR}_\xi \tag{9}$$

where CVaR_{z_0} is the risk measure imposed by the recourse costs to handle price uncertainty.

$$\text{CVaR}_{z_0} = \text{VaR}_1 + \frac{1}{1-\alpha} \sum_s \sum_i p_s \left(c_{i,s} x_{i,s} - \text{VaR}_1 \right) \tag{10}$$

where CVaR_ξ is the risk measure imposed by the recourse costs to handle uncertainty in demands and yields.

$$\text{CVaR}_\xi = \text{VaR}_2 + \frac{1}{1+\alpha} \sum_{i \in I} \sum_{s \in S} p_s \left[\left(c_i^+ z_{i,s}^+ + c_i^- z_{i,s}^- \right) + \sum_{k \in K} \left(q_{i,j}^+ y_{i,k,s}^+ + q_{i,j}^- y_{i,k,s}^- \right) - \text{VaR}_2 \right] \tag{11}$$

Substituting equations (10) and (11) into (9), we obtain a two-stage stochastic programming model with mean–risk objective in which the risk measures are assessed by CVaR.

5. Numerical Experiments

We illustrate the risk modeling approach proposed in this paper on the numerical example taken from Khor et al. (2008) and provide major details on the implementation using GAMS/CONOPT3.

5.1. Solving Two-Stage Stochastic Program with MAD as Risk Measure
The expectation of the objective function value is given by the original objective function itself: The corresponding expression for expected profit is formulated for the 13 scenarios that has been randomly generated.

5.2. Solving Two-Stage Stochastic Program with CVaR as Risk Measure
The following is the procedure for developing a loss distribution in order to determine the value for the parameter VaR.
Step 1. The objective value of deterministic profit for each of the 13 scenarios is computed (i.e., multiplication of flowrate and the corresponding price per unit flowrate).
Step 2. The probability of each scenario is randomly generated using Monte Carlo simulation based on pseudorandom number generation.
Step 3. The computed values in Step 1 are sorted in ascending order.
Step 4. The plot of cumulative distribution function against the sorted deterministic profit values is developed to obtain a representation of the loss distribution. At confidence interval of $(1-\alpha) = 0.95$, we can read off the value of VaR from the loss distribution plot, as depicted in Figures 1 and 2, which represents the penalty for uncertainty in prices and in both demands and yields, respectively.
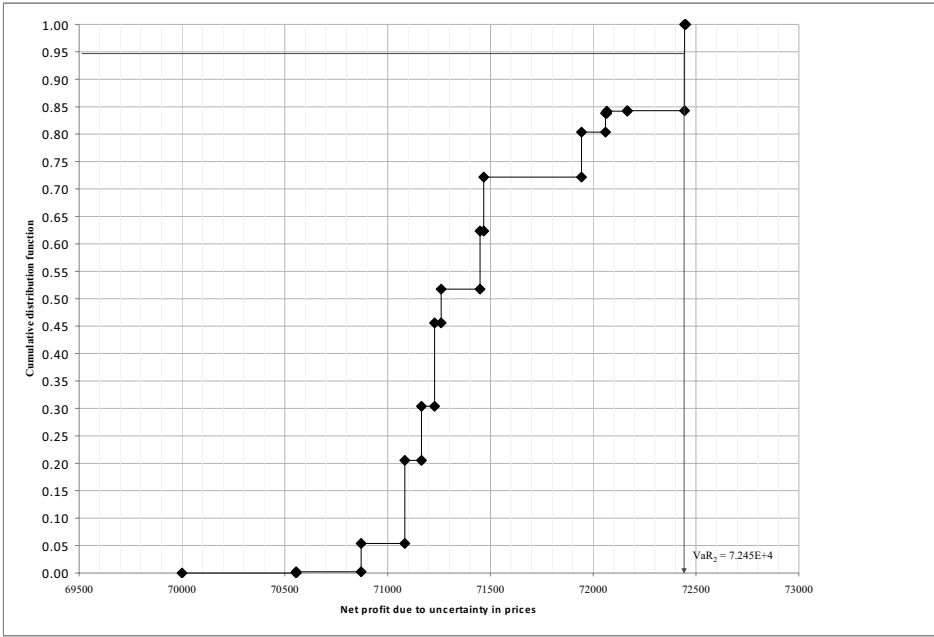
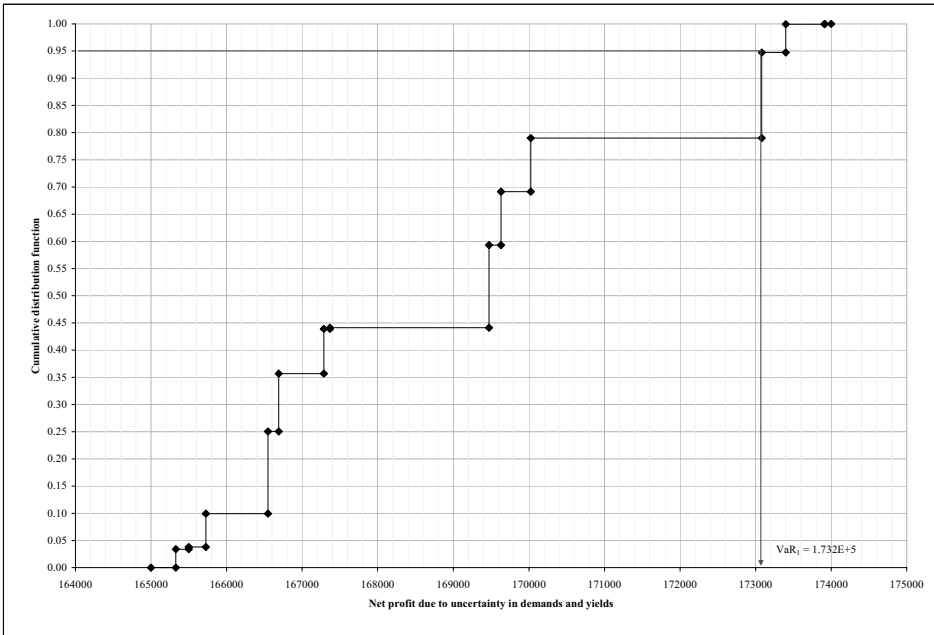Figure 1: Loss distribution to determine VaR_1

Figure 2: Loss distribution to determine VaR_2

The computational statistics and a summary of the main computational results are provided in Tables 1 and 2 , respectively.

Table 1. Summary of computational results

Monte Carlo sampling variance estimator $S(n)$	489.4
Lower bound of confidence interval H	965.3
Upper bound of confidence interval H	1237
Range of confidence interval H	271.3
Minimum number of scenarios N	13
Optimal solution for MAD-based model	\$681.95/day
Optimal solution for CvaR-based model	\$20 800.66/day

Table 2. Computational statistics of GAMS implementation for determining optimal solutions of MAD- and CVaR-based mean–risk stochastic program

Solver	GAMS/CONOPT3
Number of continuous variables	281
Number of single equations	145
CPU time/resource usage	(trivial)
Number of iterations	20 (using MAD)
	101 (using CVaR)

6. Conclusions

This work attempts to consider the use of the risk metrics of MAD and CVaR for the explicit handling of economic and operational risk management in refinery planning problems under uncertainty in prices, demands, and yields.

References

R. B. Webby, P. T. Adamson, J. Boland, P. G. Howlett, A. V. Metcalfe, and J. Piantadosi, 2007, The Mekong—applications of value at risk (VaR) and conditional value at risk (CVaR) simulation to the benefits, costs, and consequences of water resources development in a large river basin, Ecological Modeling, 201, 89–96.

C. S. Khor, A. Elkamel, K. Ponnambalam, and P. L. Douglas, 2008, Two-Stage Stochastic Programming with Fixed Recourse via Scenario Planning with Financial and Operational Risk Management for Petroleum Refinery Planning under Uncertainty, Chemical Engineering and Processing, 47, 9–10, 1744–1764.

H. Konno, and H. Yamazaki, 1991, Mean Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market. Management Science. 37, 519–531.

W. K, Mak, D. P. Morton, and R. K. Wood, 1999, Monte Carlo Bounding Techniques for Determining Solution Quality in Stochastic Programs, Operations Research Letters, 24, 47–56.

H. M. Markowitz, 1952, Portfolio selection, Journal of Finance, 7, 77–91.

R. T. Rockafellar and S. Uryasev, 2000, Optimization of conditional value-at-risk, Journal of Risk, 2, 3, 21–41.

R. T. Rockafellar and S. Uryasev, 2002, Conditional Value-at-Risk for general loss distributions, Journal of Banking and Finance, 26, 7: 1443–1471.

A. Shapiro, and T. A. Homem-de-Mello 1998, A simulation-based approach to two-stage stochastic programming with recourse, Mathematical Programming 81, 301–325.

A. Shapiro, 2000, Stochastic Programming by Monte Carlo Simulation Methods, Stochastic Programming E-Prints Series, Retrieved February 10, 2007 from the World Wide Web: http://www.isye.gatech.edu/%7Eashapiro/publications/SBOsurvey.pdf

F. You, J. M. Wassick, and I. E. Grossmann. Risk management for a global supply chain planning under uncertainty: models and algorithms [textfile]. Retrieved September 26, 2008 from the World Wide Web: http://egon.cheme.cmu.edu/Papers/RiskMgmtDow.pdf

1971

Oil Products Distribution Systems: Decomposition Approach on Pipeline and Inventory Scheduling

Susana Relvas,[a*] Ana Paula Barbosa-Póvoa,[a] Henrique A. Matos[b]

[a]*CEG-IST, UTL, Av. Rovisco Pais 1049–001 Lisboa, Portugal*
[b]*CPQ, IST, UTL, Av. Rovisco Pais 1049–001 Lisboa, Portugal*
[*]*corresponding author: susanaicr@ist.utl.pt*

Abstract

This paper addresses the problem of oil products scheduling in a system where the supply is provided by a pipeline and a tank farm that enables storage and market fulfillment. Multiproduct pipelines are complex equipments where the operational time window between product load to pipeline until product shipment for customers is wide, requiring longer scheduling horizons. However, problem complexity rises significantly with the time horizon extent and if the problem is approached through an exact optimization method such as mathematical programming the effort to obtain a solution is considerably high. This problem motivates the approach proposed within this paper consisting of an integrated framework that aims problem complexity reduction and lower computational effort. The approach combines two continuous time MILP models with different levels of detail with an iterative procedure that exchanges information between both levels. The approach is exemplified using a real world scenario of a Portuguese company that transports and distributes six oil products.

Keywords: Pipeline, Tank Farm, MILP, Decomposition.

1. Introduction

The oil supply chain is a complex dynamic system, where scheduling and planning are challenging problems to address. Besides the integration of several entities, the proper definition and representation of each activity requires problem-oriented approaches. One example is oil products' pipelines. The multiproduct nature, continuous operation and operational lag between product load and product consumption give origin to complex problems. The challenge is to solve the problem through an exact method to obtain the optimal solution with all associated constraints, in a reduced amount of time. Current published works have mainly focused on Mixed-Integer Linear Programming (MILP) models (Cafaro and Cerdá, 2008, Magatão *et al.*, 2008, Rejowski and Pinto, 2008). However, some integrated approaches were already proposed, such as the hybrid approach by Boschetto *et al.* (2008) that deals with pipeline networks. But, dealing with medium-term horizons has been a difficult task. Cafaro and Cerdá (2008) proposed a rolling horizon where some instances may be rescheduled, but each period still looks over a short-term horizon. In the present work we propose an iterative procedure between two MILP models that represent a multiproduct pipeline system through two different levels of detail in the tank farm management.

2. Problem Statement

Figure 1 resumes the operating system in study that comprises an oil products' pipeline that pumps from a refinery to a tank farm. The distribution centre is located in a strategic market position. Each tank has a fixed product service and the clients are supplied at the distribution centre with the respective products.
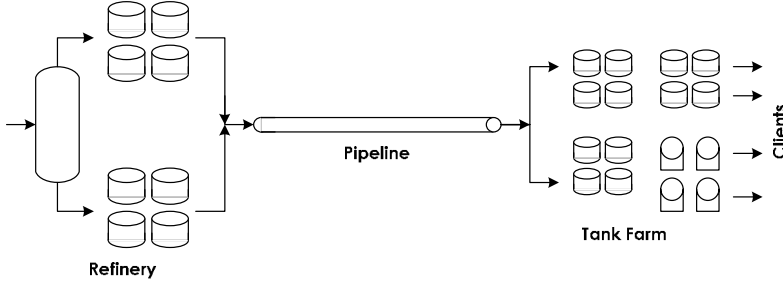

Figure 1. Multiproduct pipeline operating system

Given: a) the pipeline volume, b) maximum and minimum flowrates, c) the products to be pumped and matrix of possible sequences, d) the storage capacity by product, e) settling period by product, f) the time horizon extent, g) the maximum number of batches to be pumped, h) the initial inventory by product, i) the daily demand forecasts and k) the planned pipeline stoppages; **Determine**: 1) the pipeline schedule and 2) the inventory management. The pipeline schedule includes products' sequence, pumping flowrates, batches' volumes, timing issues and pipeline stoppages. The inventory management includes pipeline inputs, settling periods and outputs by product for customers. The problem is formulated through an MILP continuous time model and the main target is to obtain problem solutions for a time horizon compatible with the problem addressed. Short-term periods usually generate solutions where the pipeline schedule does not reflect customer requirements ahead due to the operational lag.

3. Proposed Approach

Figure 2 succinctly describes the proposed approach, which is divided in four steps:
1. sequencing heuristic;
2. sequence validation and selection through the lower detail model (ATFModel);
3. schedule generation through consecutive integer cuts (ATFModel);
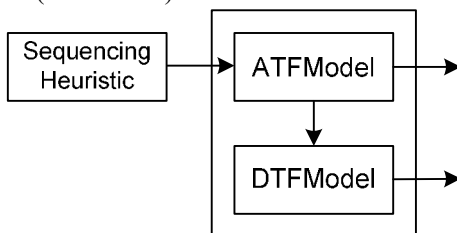4. inventory management schedule generation through a detailed model implementation (DTFModel).

Figure 2. Proposed approach: integrated MILP models with a sequencing heuristic.

The sequencing heuristic builds valid product sequences to be pumped in the given time horizon. This heuristic uses as information the market requirements and initial inventory so as to establish product priorities (Relvas *et al.*, 2009). The proposed heuristic provides inputs for the MILP models. In a first level, an Aggregated Tanks' Farm

Model (ATFModel) is run. This model considers that the tank farm can be represented in an aggregated manner on capacity: each product is stored in a single tank with a total capacity given by the summation of real tanks' capacity (Relvas *et al.*, 2006). With this model it is possible first to validate the proposed sequences and select one (or more) that returns valid results for the given time horizon. The schedule generated in this second step can be used at a higher level (e.g. for refinery planning). However an integrated procedure can be carried out, since the system considered can be represented through a more detailed formulation. A Detailed Tanks' Farm Model (DTFModel) is derived from the ATFModel that considers tanks as individual model entities. Although the model complexity and size increases, a higher level of detail is achieved for the operational decisions such as schedules for product receiving vs. available tanks and delivery schedules vs. tanks with products ready for shipment. This model addresses the complex inventory management problem at liquid products' tank farm.

The main procedure developed is based on the model integration approach. The ATFModel generates a first pipeline schedule to feed the DTFModel for a further generation of a feasible inventory solution. Although, the ATFModel already takes into account a rough inventory management representation a valid solution from this model may return an infeasible solution at the DTFModel level. This gap is reduced developing a battery of schedules with the ATFModel to test within the DTFModel level. This set of solutions is developed through an iterative approach where at each new iteration an integer cut is added. The integer cut eliminates from the current solution space a batch volume sequence with a certain degree of similarity. This is explained below: given the binary variable $y_{i,p}$ that assigns a product p to a batch i and defined through the sequencing heuristic, it can be disaggregated in the related binary variable $ls_{i,p,lt}$ that assigns a valid volume lt to batch i of product p. Therefore, an integer cut in the form of equation (1) is added at each model iteration.

$$\sum_{i \in I} \sum_{p \in P} \sum_{lt \in LT} ls_{i,p,lt} \Bigg|_{\left(ls_{i,p,lt} \right)_{itern-1}=1} \leq |I| - m, \quad \forall n \in N \tag{1}$$

where $n \in N$ represents the set of iterations, $|I|$ corresponds to the cardinality of the set of batches I and m represents a degree of diversity between consecutive batch volumes' sequences given by $ls_{i,p,lt}$. The model uses as stopping criteria either a limit on iterations reached or an infeasible solution found. The solution found in each iteration for $ls_{i,p,lt}$ is given as fixed data to the DTFModel. At each iteration, this model evaluates the feasibility of the fixed binary variables and the corresponding optimal solution with the complete pipeline schedule and inventory management.

4. Framework Implementation and Results

The proposed approach was implemented using as motivating example the real world scenario of Companhia Logística de Combustíveis (CLC), a Portuguese company that operates a multiproduct pipeline (single origin and single destination) and distributes six different oil products. At the ATFModel level, there will be six storage tanks, while at the DTFModel level there are 29 tanks, each with a fixed product service. Two case studies were considered: E1 with a one-week scenario and E2 with a two-week scenario with data information from September 2008.

The models were implemented in GAMS 22.6 and solved with CPLEX 11.0, on a Pentium D820 with 2 GHz RAM.

The extent of the set of iterations considered was 50 and the diversity parameter m was considered to be equal to 1 in E1 and 4 in E2. As stopping criteria it was used the optimal solution or, in case of E2/DTFModel, a computational time limit of 600 s.

The sequencing heuristic returned a set of 6 sequences per case study. E1 sequences had a number of batches between 6 and 11 while E2 had sequences with a number of batches between 11 and 16. Theses sequences were implemented in the ATFModel and compared through a series of operational and model performance criteria so as to select the final sequence. The E1 pumping sequence has 7 batches and for E2 has 14 batches.

Table 1 details the model size for each case study and for each model. The model size clearly increases with the time horizon extent, with the number of batches pumped and with the level of detail of the model.

Table 1. Model size by case study and model formulation

	E1		E2	
	ATFModel	DTFModel	ATFModel	DTFModel
Constraints	1867[*]	5049	6683[*]	11887
Continuous Variables	863	1240	2921	3254
Binary Variables	285	1027	964	2226

[*]the referred number of constraints corresponds to the problem before implementing the iterative procedure and, this means, before adding integer cuts.

Figure 2 represents the evolution of the objective function value with the respective iteration number. The used objective function optimizes operational variables such as minimization of the pumping flowrate and maximization (using a negative sign) of the total pipeline usage.
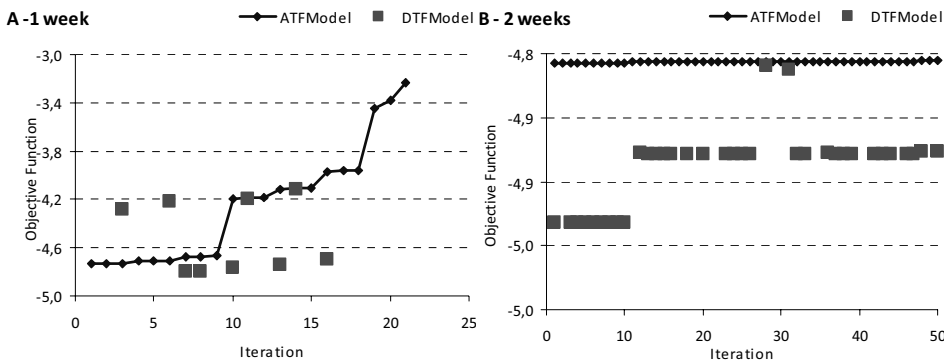

Figure 2 – Evolution of the objective function value with the iteration number in the integrated strategy

It can be observed that for E1 (Figure 2a) with only 21 iterations returned solutions whereas for case study E2 all iterations returned solutions. In terms of objective function degradation, E1 has a visible degradation trend while in E2 the trend is very smooth. It would be necessary to have either a higher diversity factor or a higher number of iterations to explore all the possible alternatives. This result emphasizes the combinatorial nature of the model. However, it can be seen that no continuous line is returned when the DTFModel is run. For a feasible ATFModel solution it is possible to have an infeasible solution at the DTFModel. This occurred for case study E1 while for

case study E2 the iterations with no solution correspond to instances where a feasible solution was not found within the maximum computational time of 600 s. In the case of E1, the solutions returned for the DTFModel are either similar, better or worse than the solutions returned from the ATFModel. On the contrary, the solutions found in the case of DTFModel in E2 are all better than for the ATFModel.

Table 2. Computational results

	E1		E2	
	ATFModel	DTFModel	ATFModel	DTFModel
Optimal Solutions	21	10	50	33
Total Number of Feasible Solutions	21	10	50	35
Total CPU effort (s)	7.8[*]	55.6/41.6[*]	4100.4	20350.6/10117.7[*]
Medium CPU effort (s)	0.37	4.62	82.01	316.18[*]
CPU effort (s) (no integration)	-	7.4	-	7225.2

[*]considering only optimal solutions

In terms of computational effort, table 2 presents the results summary. E1 used less than 10 seconds to run all iterations at the ATFModel level and less than 60 seconds to run all iterations at the DTFModel level. For comparison, it was run the DTFModel with no information provided from the ATFModel and the optimal solution was found in 7.4 s. When the time horizon expands to the double (from E1 to E2), the DTFModel without integration requires about 2 hours of computation time to prove optimality. For this scenario, however, the relative gap was 0.02%. The set of iterations was run in 4100 s for the ATFModel (in average each iteration takes 82 s) but at the DTFModel, 35 solutions were found after a total computational time of 20350 s (5.65 h).

Finally, table 3 presents the operational results obtained for each scenario and each model, as well as the correspondent schedule data developed by CLC's schedulers. It also has information about the iterations where the best objective function value was found for each model. While for E1 there is no correspondence between the iteration where the best solutions are found, in the case of E2 there is.

Table 3. Operational results

	E1			E2		
	ATFModel	DTFModel	CLC	ATFModel	DTFModel	CLC
Iteration with best FO	1-2	7-8	-	1-10	1;3-10	-
Medium flowrate (vu/h)[*]	469.9	470.4	519.0	465.2	466.2	494.6
ΔInventory (vu)[*]	+2350	+450	+4769	+265	+265	+11105
Pipeline usage (%)	95.9	93.4	89.6	93.8	93.7	94.8

[*]vu stands for volumetric units

Some operational indicators were selected for comparison. The medium flowrate is calculated for the period when the pipeline is pumping, taking into account the flowrate employed for each batch. The ΔInventory represents the difference between total pipeline inputs and total outputs for the local market. Finally, the pipeline usage is obtained by the ratio of total pumping time and total horizon time.

If the operational indicators are compared for each case study, there is a negligible degradation between the runs of the ATFModel and the DTFModel. However, when comparing the schedule developed by CLC's schedulers, it can be seen that the medium flowrate is higher and, at the end, it is transported a considerable amount of volume than it is consumed. However, when looking in depth over the flowrates by batch, the models propose flowrates no higher than 470 vu/h in E1 and no higher than 490 vu/h in E2, having as lower bound the value of 450 vu/h. In the schedule proposed by CLC's schedulers, flowrates vary between 450 and 700 vu/h. In terms of pipeline usage, the values are similar, except a lower pipeline usage proposed by CLC's schedulers in the first week.

5. Conclusions and Future Work

This paper proposes an approach to address short to medium term horizons in multiproduct pipeline scheduling problems with inventory management. The approach uses an iterative procedure combined with sequential valid integer cuts that explore the solution space at a low detail level model to provide initial data for a higher detail model. With this strategy it is possible to tackle longer pipeline time horizons and it is possible to explore two levels of detail and a set of solutions rather than having a single time consuming solution. This approach meets real world solutions requirements: lower computational effort and a valid set of solutions for comparison and selection. Additionally, when comparing with a real world scenario where schedulers use trial and error methods based in spreadsheets, the improvements are significant.

As future work, the authors propose to study the influence of the diversity degree between consecutive iterations, a selection of number of iterations criteria as well as the exchange of more information between models, such as pumping time windows per batch. The authors also propose to analyze how the approach behaves for higher time horizons, up to one month of time span, in order to provide pipeline requirements for the refinery monthly production planning.

6. Acknowledgments

References

S.N. Boschetto, L.C. Felizari, L. Yamamoto, L. Magatão, S.L. Stebel, F. Neves Jr., L.V.R. Arruda, R. Lüders, P.C. Ribas, L.F.J. Bernardo, 2008, An Integrated Framework for Operational Scheduling of a Real-World Pipeline Network, In: Bertrand Braunschweig; Xavier Joulia. (Editors), Computer Aided Process Engineering, Elsevier , vol 25, 259;

D.C. Cafaro, J. Cerdá, 2008, Dynamic scheduling of multiproduct pipelines with multiple delivery due dates, Comp. & Chem. Eng., 32, 728;

L. Magatão, L.V.R. Arruda, F. Neves, Jr, 2004, A mixed integer programming approach for scheduling commodities in a pipeline, Comp. & Chem. Eng., 28, 171;

R. Rejowski Jr, J.M. Pinto, 2008, A continuous time representation for scheduling of pipeline systems with pumping yield rate constraints, Comp. & Chem. Eng., 32, 1042;

S. Relvas, H.A. Matos, A.P.F.D. Barbosa-Póvoa, J. Fialho, A.S. Pinheiro, 2006, Pipeline scheduling and inventory management of a multiproduct distribution oil system, Ind. Eng. Chem. Res., 45, 7841

S. Relvas, A.P.F.D. Barbosa-Póvoa, H.A. Matos, 2009, Heuristic batch sequencing on a multiproduct oil distribution system, Comp. & Chem. Eng., 33, 712.

1977

Efficient Bulk Maritime Logistics for the Supply and Delivery of Multiple Chemicals

Jie Li,[a] I. A. Karimi,[a] Rajagopalan Srinivasan[a,b]

[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore, 117576*
[b]*Process Sciences and Modeling, Institute of Chemical and Engineering Science, 1 Pesek Road, Jurong Island, Singapore 627833*

Abstract

In this paper, we develop a novel unit-slot mixed integer linear programming (MILP) continuous-time model for a ship planning problem, in which a heterogeneous fleet of ships is used to supply and deliver multiple chemical products between production and consumption sites. We incorporate many real-life features such as multiple pickups and deliveries of multiple products using ships with multiple dedicated compartments, variable load/discharge quantities, multiple jetties, and fixed setup time for each product, etc. Furthermore, our model ensures product inventories at any time to meet their minimum and maximum capacities. Four examples are solved to illustrate the superiority and efficiency of our proposed formulation.
Keywords: IPDP, Mixed-integer linear programming (MILP), bulk product transportation, unit slot.

1. Introduction

Maritime transportation is very critical to the world economy, because approximately 90% by volume and 70% by value of all goods are transported worldwide by sea (Psaraftis, 1998). Moreover, approximately 80% of all maritime ton-miles is via bulk transportation, which offers the lowest cost per ton-mile. Therefore, operational efficiency of maritime transportation can reduce final product costs significantly.

In this paper, we address a ship planning problem, in which a heterogeneous fleet of ships with dedicated compartments is used to transport multiple chemicals between production and consumption sites. This is done while ensuring that adequate inventory levels of all chemicals are maintained at all consumption sites. This problem is termed as Inventory Pickup and Delivery Problem (IPDP), which has been addressed previously by Christiansen (1999), Al-Khayyal and Hwang (2007) and Li et al. (2008). However, they all have some limitations. For instance, the model of Christiansen (1999) lacked many realistic operation features such as multiple dedicated compartments, multiple chemicals, variable production/consumption rates, multiple jetties, setup times, etc. Al-Khayyal and Hwang (2007) did not allow several ships to load/discharge the same product at each site simultaneously. Moreover, they could not ensure that product inventories were maintained between allowable limits at any time. Lastly, Li et al. (2008) did not incorporate setup times and were unable to solve large-scale problems.

In this paper, we develop a novel unit-slot mixed integer linear programming (MILP) continuous-time formulation for this chemical logistics problem and incorporate many real-life features such as multiple pickups and deliveries of multiple products using ships with multiple dedicated compartments, variable load/discharge quantities, product load/discharge sequences, variable production/consumption rates, multiple jetties, and

fixed setup time for each product. Most importantly, our model ensures product inventories at any time to meet their minimum and maximum capacities and allows several ships to load/discharge the same product at each site simultaneously. Furthermore, one site can be visited several times by a ship. Four examples are solved to illustrate the efficiency of our proposed formulation.

2. Problem Statement

Most multinational companies have plants distributed all over the world. Some of these may produce chemical products that are raw materials to other plants, while others may simply consume products. Thus, a plant can be a production site, consumption site, or both. If a site does not have sufficient inventory of a product, then it may have to purchase from external suppliers in the local market. Thus, we define two types of sites. Internal site belongs to a company, while an external site does not. We consider inventory at internal sites only. Consider I (i = 1, 2, 3, …, I) sites, B (b = 1, 2, 3, …, B) jetties, and M (m = 1, 2, …, M) products in exactly M dedicated product tanks. A heterogeneous fleet of ships V (v = 1, 2, …, V) with multiple dedicated compartments is used to load a product from a production site and discharge it at a consumption site to ensure enough inventory at each site. Figure 1 shows an example transport scenario. With this, the problem can be stated as follows.
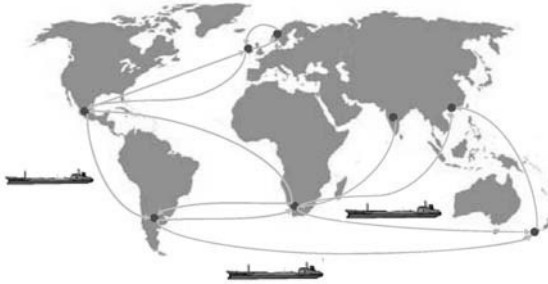
Figure 1 An example transport scenario

Given:
1. V ships, products that they can carry, minimum and maximum capacities for these products, initial position, initial products and their amounts, load and discharge rates.
2. I Sites, products that they can produce or consume, jetties, initial holdups, minimum and maximum capacities of products, limits on production or consumption rates.
3. Ship travel times between sites, and setup times for switching from a product to another.
4. Ship travel costs between sites, fixed cost for product load or discharge at each site.
Decide:
1. Which ship should load or discharge which product at which site, at which time, and in what amounts
2. The inventory profile of each product at each site.

3. MILP Formulation

We divide the scheduling horizon [0, H] into K (k = 1, 2, …, K) contiguous slots (Figure 2) and denote the time before the horizon start by slot zero (k = 0). We define ships, jetties, and product tanks as various units. Each unit has K slots, which are not synchronized (Lim & Karimi, 2003) across the units. In other words, they are unit slots (Liu & Karimi, 2008), where the start/end times and slot lengths of a given slot k need

not be the same across all units. We let T_{qk} ($k = 0, 1, 2, \ldots, K$; $T_{q0} \geq 0$, $T_{qK} \leq H$) denote the end time of slot k on unit q, where q becomes v for a ship, becomes b for a jetty, and becomes m for a product tank. Slot k on unit q starts at $T_{q(k-1)}$ and ends at T_{qk}. Since the slots are asynchronous, T_{vk}, T_{bk}, and T_{imk} may vary with units. Thus,

$$T_{qk} \geq T_{q(k-1)} \qquad\qquad 1 \leq k \leq K \qquad\qquad (1)$$

Figure 2 Schematic of unit slots design

In the following, $\mathbf{B}_i$ is the set of jetties that belong to site i, $\mathbf{M}_i$ is the set of products that can be produced or consumed at site i, $\mathbf{M}_v$ is the set of products that can be carried by ship v, $\mathbf{V}_i$ is the set of ships that can visit site i, $\mathbf{I}_v$ is the set of sites that ship v can visit, and $\mathbf{IT}_i$ is the set of internal sites.

3.1. Ship Load or Discharge Operations

We use two dummy sites ($i = 0, I + 1$) to denote the initial and final positions of a ship respectively. Thus, we have $I+2$ sites including I real site ($i = 1, 2, \ldots, I$) and two dummy ($i = 0, I + 1$) sites. Suppose a ship v is loading or discharging products in its own slot k, whereas a product tank m is receiving from or feeding this v in its own slot k'. Three scenarios are possible: $k' < k$, $k' = k$, and $k' > k$. For $k' < k$, we can simply introduce additional slots on product tank i to make $k' = k$. For $k' > k$, we can do the same on ship v. In other words, with no loss of generality, we can demand that if a ship v is loading or discharging products from or to a product tank m at any time, then the unit slots corresponding to that time on both ship v and product tank m must have the same index. The same holds true for a ship and a jetty. Thus, we now define binary variables X_{ikv}, Y_{mkv} and y_{bkv} and 0-1 continuous variables ye_{mkv} and Z_{ijkv}.

$$X_{ikv} = \begin{cases} 1 & \text{If ship } v \text{ visits site } i \text{ during slot } k \\ 0 & \text{Otherwise} \end{cases} \qquad 0 \leq i \leq I+1, 0 \leq k \leq K$$

$$Y_{mkv} = \begin{cases} 1 & \text{If ship } v \text{ loads/discharges product } m \text{ during slot } k \\ 0 & \text{Otherwise} \end{cases} \qquad 0 < k \leq K$$

$$y_{bkv} = \begin{cases} 1 & \text{If ship } v \text{ loads/discharges via jetty } b \text{ in slot } k \\ 0 & \text{Otherwise} \end{cases} \qquad 0 < k \leq K$$

$$ye_{mkv} = \begin{cases} 1 & \text{If ship } v \text{ ends load/discharge product } m \text{ at the end of slot } k \\ 0 & \text{Otherwise} \end{cases} \qquad 0 < k < K$$

$$Z_{ijkv} = \begin{cases} 1 & \text{If ship } v \text{ visits site } i \text{ in slot } k \text{ and } j \text{ in } k+1 \\ 0 & \text{Otherwise} \end{cases} \qquad 0 \leq i, j \leq I+1, 0 \leq k < K$$

We treat X_{ikv} ($i = 0, 1, 2, 3, \ldots, I$) as binary and $X_{(I+1)kv}$ as 0-1 continuous variable. Thus, at any time, a ship v visits exactly one site,

$$\sum_{i \in \mathbf{I}_v} X_{ikv} = 1 \qquad\qquad 0 < k \leq K \qquad\qquad (2)$$

At most one ship can load or discharge through one jetty b in site i at a time. Then,

$$\sum_{v \in \mathbf{V}_i} y_{bkv} \leq 1 \qquad\qquad b \in \mathbf{B}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \qquad\qquad (3)$$

3.2. Slot Timings on Product Tanks

When using unit slots in the presence of shared resources such as inventories, the main challenge is to relate the timings of different units that share the same resource. The flow in/out of a resource must be ordered chronologically, so that we can get a correct resource profile. If a ship v is loading or discharging product m during slot k, then the start (end) of a slot k on a product tank m must precede (succeed) the start (end) of slot k on the ship.

$$T_{im(k-1)} \leq T_{v(k-1)} + H(2 - X_{ikv} - Y_{mkv})$$

$$i \in \mathbf{IT}_i, \ v \in \mathbf{V}_i, \ m \in \mathbf{M}_v \ni \mathbf{M}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \qquad (4a)$$

$$T_{imk} \geq T_{v(k-1)} + RU_{imkv} - H(2 - X_{ikv} - Y_{mkv})$$

$$i \in \mathbf{IT}_i, \ v \in \mathbf{V}_i, \ m \in \mathbf{M}_v \ni \mathbf{M}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \qquad (4b)$$

For checking the inventory at the end of each slot, we define an intermediate point t_{imk} between $T_{im(k-1)}$ and T_{imk}. Then, we demand that the end of each loading or discharging in slot k must match with this point on the tank by using the following.

$$T_{v(k-1)} + RU_{imkv} \leq t_{imk} + H(2 - X_{ikv} - Y_{mkv})$$

$$i \in \mathbf{IT}_i, \ v \in \mathbf{V}_i, \ m \in \mathbf{M}_v \ni \mathbf{M}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \qquad (5a)$$

$$T_{v(k-1)} + RU_{imkv} \geq t_{imk} - H(2 - X_{ikv} - Y_{mkv})$$

$$i \in \mathbf{IT}_i, \ v \in \mathbf{V}_i, \ m \in \mathbf{M}_v \ni \mathbf{M}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \qquad (5b)$$

3.3. Site Inventory

The timing constraints of section 3.2 enable us to write the following inventory balances for product tanks.

$$Iv_{imk} = Iv_{im(k-1)} - JJ_{im} \sum_{v \in \mathbf{V}_i} qv_{imkv} + JJ_{im} SQ1_{imk} \qquad i \in \mathbf{IT}_i, \ m \in \mathbf{M}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \quad (6a)$$

$$Iv_{im}^L \leq Iv_{im(k-1)} - JJ_{im} \sum_{v \in \mathbf{V}_i} qv_{imkv} + JJ_{im} SQ2_{imk} \leq Iv_{im}^U$$

$$i \in \mathbf{IT}_i, \ m \in \mathbf{M}_i, \ 1 \leq i \leq I, \ 0 < k \leq K \quad (6b)$$

where, Iv_{imk} ($Iv_{im}^L \leq Iv_{imk} \leq Iv_{im}^U$) is the inventory of product m at site i at the end of slot k. $SQ1_{imk}$ is the amount of product m produced/consumed at site i in slot k, and $SQ2_{imk}$ is the amount of product m at site i produced/consumed from $T_{im(k-1)}$ to t_{imk}. JJ_{im} is 1 if site i produces m, otherwise -1. The expressions for $SQ1_{imk}$ and $SQ2_{imk}$ are not given here.

3.4. Objective Function

Our objective is to minimize total operating cost over the planning horizon, which consists of traveling and load/discharge costs.

$$\text{Min } TC = \sum_{v=1}^{V} \sum_{i \in \mathbf{I}_v} \sum_{j \in \mathbf{I}_v \ni j \neq i} \sum_{k=0}^{K} C_{ijv} Z_{ijkv} + \sum_{v=1}^{V} \sum_{m \in \mathbf{M}_v} \sum_{k=1}^{K} C_m ye_{mkv} + \sum_{v=1}^{V} \sum_{m \in \mathbf{M}_v} C_m Y_{mKv} \qquad (7)$$

where, C_{ijv} is the travelling cost of ship v from sites i to j. C_m is the load/discharge cost of product m.

So far, our complete model comprises eqs. 1-7 and other equations such as slot timings on jetties, inventory balance at the end of the planning horizon, etc. that are given above.

4. Results and Discussions

We solve four examples to evaluate our proposed formulation. These examples vary from ships, sites, products, jetties, etc. and are solved using CPLEX 10.0.1/GAMS 22.2 on a Dell workstation PWS690 (Intel® Xeon™ CPU 3.00 GHZ, 16 GB memory) running Windows XP. Table 1 shows solution statistics for Examples 1-4. Note that the models of Christiansen (1999) and Al-Khayyal and Hwang (2007) need the estimated arrival times of each site. We use the same estimated arrival times for all sites. To choose the best arrival times, we start from a small number of arrival times, and then increase by 1 until the objective function does not change. This approach is the same as that of choosing the best slot.

Table 1 Solution statistics of various models for Examples 1-4

Example	Model	Discrete Variables	Continuous Variables	Constraints	CPU Time (Seconds)	MILP Solution	Relative Gap (%)
1	Ours	40	359	592	0.73	7	0
	Li et al. (2008)	120	831	1050	1.77	7	0
	Al-Khayyal and Hwang (2007)	158	161	719	N/A	N/A	N/A
2	Ours	35	485	642	0.78	32.5	0
	Li et al. (2008)	106	967	905	1.84	32.5	0
	Christiansen (1999)	336	177	1256	3.58	32.5	0
3	Ours	415	2659	3369	706.8	22	0
	Li et al. (2008)	597	4861	4301	86400	28	14.03
4	Ours	105	1092	1101	136.3	50.4	0
	Christiansen (1999)	952	341	3317	354.3	50.4	0

N/A: The obtained solution is infeasible
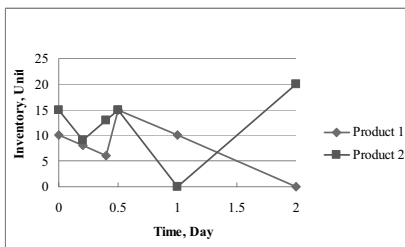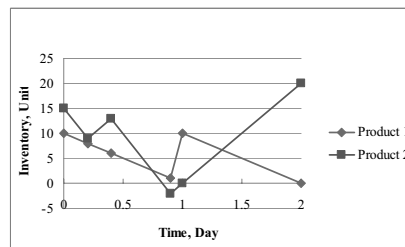
Figure 3 Inventory profiles of products 1 and 2 at site 1 for Example 1 with different product load/discharge sequences: (a) first load/discharge product 1 (b) first load/discharge product 2

Example 1

This example involves two ships with two products, and three sites, each of which has two jetties. While site 1 consumes product 1 and produces product 2, sites 2 and 3 both consume product 2 and produce product 1. The capacities of products 1 and 2 at sites 1 are [0, 12 unit] and [0, 21 unit]. Figure 3 gives the optimal inventory profiles of products 1 and 2 at site 1 from the model of Al-Khayyal and Hwang (2007). While the inventory of product 1 at 0.5 day is 15 unit in case (a), which violates its maximum capacity (12 unit), that of product 2 at 0.9 day is -2 unit in case (b), which also violates its minimum capacity (0). Hence, the model of Al-Khayyal and Hwang (2007) cannot guarantee the inventories of products to meet their minimum and maximum capacities

throughout the entire planning horizon. The model of Christiansen (1999) cannot be used to solve this example because it is restricted to single product. The model of Li et al. (2008) and our model obtain the optimal solution of 7. While the model of Li et al. (2008) needs 1.77 CPU s, our model needs only 0.73 CPU s. No violation on inventory capacities occurs through the entire horizon with our model.

Example 2

This Example is taken from Li et al. (2008) involving two ships, five sites and one single product with 16-day planning horizon. We obtain the optimal solution of 32.5 within 0.78 CPU Seconds, while the model of Li et al. (2008) needs 1.84 CPU seconds and the model of Christiansen (1999) needs 3.58 CPU s. When only one product is involved, the model of Al-Khayyal and Hwang (2007) is the same as that of Christiansen (1999).

Examples 3-4

Example 3 consists of five ships, eight sites, and two products. The planning horizon is 80 days. Example 4 involves five ships, five sites, and one product with 60-day planning horizon. For Example 3, our model obtains the optimal solution of 22 within 706.8 CPU s. However, the model of Li et al. (2008) gets suboptimal solution of 28 with 14.03% relative gap after 24 hours. Since the model of Christiansen (1999) is only for single product and the model of Al-Khayyal and Hwang (2007) may violate inventory capacities when two or more products are involved, they are not used to solve this example. For Example 4, our model obtains the optimal solution of 50.4 within 136.3 CPU s. However, the model of Christiansen (1999) needs 354.3 CPU s to obtain the optimal solution.

5. Conclusion

In this paper, we developed an efficient unit-slot MILP formulation for a ship planning problem. We incorporated many more real-life features compared to the models of Christiansen (1999) and Al-Khayyal (2007). Furthermore, our model ensured product inventories to meet their minimum and maximum capacities at any time during the planning horizon. On 4 test problems of varying sizes, our model performed better than those of Li et al. (2008) and Christiansen (1999).

References

F. Al-Khayyal, S. J. Hwang, 2007, Inventory constrained maritime routing and scheduling for multi-commodity liquid bulk, Part I: Applications and model, *European J. of Operational Research*, 176, 106-130.

M. Christiansen, 1999, Decomposition of a combined inventory and time constrained ship routing problem, *Transportation Science*, 33, 1, 3-16.

J. Li, I. A. Karimi, R. Srinivasan, 2008, Supply and distribution of multiple products via bulk maritime logistics, *5th Foundations of Computer-Aided Process Operations- FOCAPO*, Cambridge, Massachusetts, USA, June 29- July 2.

M. F. Lim, I. A. Karimi, 2003, Resource-constrained scheduling of parallel production lines using asynchronous slots, *Ind. Eng. Chem. Res.*, 42, 6832-6842.

Y. Liu, I. A. Karimi, 2008, Scheduling multistage batch plants with parallel units and no interstage storage, *Comput Chem Eng.*, 32, 671-693.

H. N., Psaraftis, 1999, Forward to focused issue on maritime transportation, *Transportation Science*, 33, 1-2.

MINLP Model and Algorithms for Optimal Design of Large-Scale Supply Chain with Multi-Echelon Inventory and Risk Pooling Under Demand Uncertainty

Fengqi You, Ignacio E. Grossmann

Department of Chemical Engineeirng, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

We address the optimal design of a multi-echelon supply chain and the associated inventory systems in the presence of uncertain customer demands. By using the guaranteed service approach to model the multi-echelon stochastic inventory system, we develop an optimization model for simultaneously optimizing the transportation, inventory and network structure of a multi-echelon supply chain. We formulate this problem as an MINLP with a nonconvex objective function including bilinear, trilinear and square root terms. By exploiting the properties of the basic model, we reformulate the problem as a separable concave minimization program. A spatial decomposition algorithm based on Lagrangean relaxation and piecewise linear approximation is proposed to obtain near global optimal solutions with reasonable computational expense. Examples for industrial gas supply chains with up to 5 plants, 50 potential distribution centers and 100 markets are presented.

Keywords: Supply Chain, Safety Stock, Risk-pooling, Uncertainty, MINLP.

1. Introduction

Due to the increasing pressure for remaining competitive in the global market place, optimizing inventories across the supply chain has become a major challenge for the process industries to reduce costs and to improve the customer service (Grossmann, 2005). This challenge requires integrating inventory management with supply chain network design, so that decisions on the locations to stock the inventory and the associated amount of inventory in each stocking location can be determined simultaneously for lower costs and higher customer service level. However, the integration is usually nontrivial for multi-echelon supply chains and their associated inventory systems in the presence of uncertain customer demands (Zipkin, 2000).

The objective of this work is to develop optimization models and solution algorithms to address the problem of joint multi-echelon supply chain network design and inventory management. By using the guaranteed service approach to model the multi-echelon inventory system (Graves & Willems, 2005) we capture the stochastic nature of the problem, and develop an equivalent deterministic optimization model. The model determines the supply chain design decisions such as the locations of distribution centers (DCs), assignments of markets to DCs, assignments of DCs to plants, shipment levels from plants to the DCs and from DCs to customers, and inventory decisions such as pipeline inventory and safety stock in each node of the supply chain network. The

model also captures risk-pooling effects (Eppen, 1979) by consolidating the safety stock inventory of downstream nodes to the upstream nodes in the multi-echelon supply chain. The model is first formulated as a mixed-integer nonlinear program (MINLP) with a nonconvex objective function, and then reformulated as a separable concave minimization program. To solve the problem efficiently, a decomposition algorithm based on Lagrangean relaxation and piece-wise linear approximation is developed to obtain near global optimal solutions within 1% optimality gap with modest CPU times. Examples are presented to illustrate the application of the model and its performance.

2. Problem Statement

We are given a potential supply chain consisting of a set of plants (or suppliers), a number of candidate sites for distribution centers, and a set of customer demand zones whose inventory costs should be taken into account. The market can represent a local distributor, a regional warehouse, a dealer, a retailer, or a wholesaler. Alternatively, one might view the customer demand as the aggregation of a group of customers operated with vendor managed inventory, which is a common business model in the industrial gases industry and some chemical companies.

In the given potential supply chain, the locations of the plants, potential distribution centers and markets are known and the distances between them are given. The investment costs for installing DCs are expressed by a cost function with fixed charges. Each market k has an uncorrelated normally distributed demand with mean μ and variance σ in each unit of time. Single sourcing restriction, which is common in the industries of specialty chemical and industrial gases, is employed for the distribution from plants to DCs and from DCs to markets. That is, each DC is only served by one plant, and each market is only assigned to one DC to satisfy the demand. Linear transportation costs are incurred for shipments from plant i to distribution center j with unit cost, and from distribution center j to market k with unit cost. The corresponding deterministic order processing times of DCs and market that includes the material handling time, transportation time and inventory review period. The service time of each plant, and the maximum service time of each markets are known. We are also given the safety stock factor for DCs and markets, and, which correspond to the standard normal deviate of the maximum amount of demand that the node will satisfy from its safety stock. A common review period is used for the control of inventory in each node. Inventory costs are incurred at DCs and markets, and consist of pipeline inventory and safety stock, of which the unit costs are given. The objective is to determine how many distribution centers (DCs) to install, where to locate them, which plants to serve each DC and which DCs to serve each market, how long should each DC quote its service time, and what level of safety stock to maintain at each DC and market so as to minimize the total installation, transportation, and inventory costs.

3. Model Formulation

The joint multi-echelon supply chain design and inventory management model is a mixed-integer nonlinear program (MINLP) that deals with the supply chain network design for a given product, and considers its multi-echelon inventory management.

3.1. Objective Function
The objective function of this model (the total supply chain design cost) is given by,

$$\min: \sum_{j \in J} f_j Y_j + \sum_{j \in J}\left(g_j \sum_{k \in K} \chi Z_{jk} \mu_k \right) + \sum_{i \in I} \sum_{j \in J}\left(c1_{ij} X_{ij} \sum_{k \in K} \chi Z_{jk} \mu_k \right) + \sum_{j \in J} \sum_{k \in K} c2_{jk} \chi Z_{jk} \mu_k$$
$$+ \sum_{i \in I} \sum_{j \in J}\left(\theta1_j t1_{ij} X_{ij} \sum_{k \in K} Z_{jk} \mu_k \right) + \sum_{j \in J} \sum_{k \in K} \theta2_k t2_{jk} Z_{jk} \mu_k + \sum_{j \in J} \lambda1_j h1_j \sqrt{N_j} \cdot \sqrt{\sum_{k \in K} \sigma_k^2 Z_{jk}} + \sum_{k \in K} \lambda2_k h2_k \cdot \sigma_k \sqrt{L_k} \tag{1}$$

which includes the following items:

DC Installation Cost: The cost of installing a DC in candidate location j is expressed by a fixed-charge cost model that captures the economies of scale in the investment. The annual expected demand of DC j is ($\sum_{k \in K} \chi Z_{jk} \mu_k$), which equals to the annual mean demand of all the markets served by DC j. Hence, the cost of installing DC j consists of fixed cost f_j and variable cost ($g_j \sum_{k \in K} \chi Z_{jk} \mu_k$), which is the product of variable cost coefficient and the expected demand of this DC in one year. Thus, the total installation cost of all the DCs is given by $\sum_{j \in J} f_j Y_j + \sum_{j \in J}\left(g_j \sum_{k \in K} \chi Z_{jk} \mu_k \right)$.

Transportation costs from plants to DCs and from DCs to markets: The product of the annual mean demand of DC j and the unit transportation cost ($\sum_{i \in I} c1_{ij} X_{ij}$) between DC j and the plant that serves it yields the annual plant to DC transportation cost, $\sum_{i \in I} \sum_{j \in J}\left(c1_{ij} X_{ij} \sum_{k \in K} \chi Z_{jk} \mu_k \right)$. Similarly, the product of yearly expected mean demand of market k ($\chi \mu_k$) and the unit transportation cost ($\sum_{j \in J} c2_{jk} Z_{jk}$) between market k and the DC that serves it, yields the corresponding transportation cost, $\sum_{j \in J} \sum_{k \in K} c2_{jk} \chi Z_{jk} \mu_k$.

Pipeline inventory costs in DCs and markets: Based on Little's law, the pipeline inventory PI_j of DC j equals to the product of its daily mean demand ($\sum_{k \in K} Z_{jk} \mu_k$) and its order processing time ($\sum_{i \in I} t1_{ij} X_{ij}$), which is in terms of days. Thus, the annual total pipeline inventory cost of all the DCs is given by, $\sum_{i \in I} \sum_{j \in J}\left(\theta1_j t1_{ij} X_{ij} \sum_{k \in K} Z_{jk} \mu_k \right)$, where $\theta1_j$ is the annual unit pipeline inventory cost of DC j. Similarly, the total annual pipeline inventory cost of all the markets is given by, $\sum_{j \in J} \sum_{k \in K} \theta2_k t2_{jk} Z_{jk} \mu_k$, where $\theta2_k$ is the annual unit pipeline inventory cost of market k.

Safety stock costs in DCs and markets: The demand at market k follows a given normal distribution with mean μ_k and variance σ_k^2. Due to the risk-pooling effect (Eppen, 1979), the demand over the net lead time (N_j) at DC j is also normally distributed with a mean of $N_j \sum_{k \in J_k} \mu_k$ and a variance of $N_j \sum_{k \in J_k} \sigma_k^2$, where J_k is the set of markets k assigned to DC j. Thus, the safety stock required in the DC at candidate location j with a safety stock factor $\lambda1_j$ is $\lambda1_j \sqrt{N_j} \cdot \sqrt{\sum_{k \in K} \sigma_k^2 Z_{jk}}$. Considering the annual inventory holding cost at DC j is $h1_j$, we have the annual total safety stock cost at all the DCs equals to, $\sum_{j \in J} \lambda1_j h1_j \sqrt{N_j} \cdot \sqrt{\sum_{k \in K} \sigma_k^2 Z_{jk}}$. Similarly, the demand over the net lead time of markets k (L_k) is normally distributed with a mean of $L_k \mu_k$ and a variance of $L_k \sigma_k^2$. Thus, the annual safety stock cost at all the markets is given by, $\sum_{k \in K} \lambda2_k h2_k \cdot \sigma_k \sqrt{L_k}$.

3.2. Constraints

Three constraints are used to define the network structure. The first one is that if DC j is installed, it should be served by only one plant. If it is not installed, it is not assigned to any plant. This can be modelled by,

$$\sum_{i \in I} X_{ij} = Y_j, \quad \forall j \tag{2}$$

The second constraint states that each market k is served by only one DC,

$$\sum_{j \in J} Z_{jk} = 1, \quad \forall k \tag{3}$$

The third constraint states that if a market k is served by the DC in candidate location j, the DC must exist,

$$Z_{jk} \le Y_j, \quad \forall j,k \tag{4}$$

Two constraints are used to define the net lead time of DCs and markets. The replenishment lead time of DC j should be equal to the guaranteed service time (SI_i) of plant i, which serves DC j, plus the order processing time (t_{ij}). Since each DC is served by only one plant, the replenishment lead time of DC j is given by $\sum_{i \in I}(SI_i + t_{ij}) \cdot X_{ij}$. Thus, the net lead time of DC j should be greater than its replenishment lead time minus its guaranteed service time to its successor markets, given by the linear inequality,

$$N_j \ge \sum_{i \in I}(SI_i + t1_{ij}) \cdot X_{ij} - S_j, \quad \forall j \tag{5}$$

Similarly, the net lead time of a market k is greater than its replenishment lead time minus its maximum guaranteed service time, R_k, is given by the nonlinear inequalities,

$$L_k \ge \sum_{j \in J}(S_j + t2_{jk}) \cdot Z_{jk} - R_k, \quad \forall k \tag{6}$$

Finally, all the decision variables for network structure are binary variables, and the variables for guaranteed service time and net lead time are non-negative variables.

$$X_{ij}, Y_j, Z_{jk} \in \{0,1\}, \quad \forall i,j,k \tag{7}$$

$$S_j \ge 0, \quad N_j \ge 0, \quad \forall j \tag{8}$$

$$L_k \ge 0, \quad \forall k \tag{9}$$

3.3. Reformulation

The original model is a nonconvex MINLP. To reduce the computational efforts, we use linearization techniques to reformulate the model as a separable concave minimization problem. The reformulation model (AP) has a new objective function given as follows.

$$\min \quad \sum_{j \in J} f_j Y_j + \sum_{i \in I}\sum_{j \in J}\sum_{k \in K} A_{ijk} XZ_{ijk} + \sum_{j \in J}\sum_{k \in K} B_{jk} Z_{jk} + \sum_{j \in J} q1_j \sqrt{NZV_j} + \sum_{k \in K} q2_k \sqrt{L_k} \tag{10}$$

In addition to linear constraints (2), (3), (4), (7), (8), (9), the reformulated model includes the following linear constraints for linearization the product of a binary variable and a continuous variable (Glover, 1975):

$$N_j \ge \sum_{i \in I} \bar{S}_{ij} \cdot X_{ij} - S_j, \forall j \tag{11}$$

$$XZ_{ijk} \le X_{ij}, \quad \forall i,j,k \tag{12}$$

$$XZ_{ijk} \le Z_{jk}, \quad \forall i,j,k \tag{13}$$

$$XZ_{ijk} \ge X_{ij} + Z_{jk} - 1, \forall i,j,k \tag{14}$$

$$SZ_{jk} + SZ1_{jk} = S_j, \quad \forall j,k \tag{15}$$

$$SZ_{jk} \le Z_{jk} \cdot S_j^U, \quad \forall j,k \tag{16}$$

$$SZ1_{jk} \le (1 - Z_{jk}) \cdot S_j^U, \forall j,k \tag{17}$$

$$NZ_{jk} + NZ1_{jk} = N_j , \quad \forall j,k \tag{18}$$

$$NZ_{jk} \leq Z_{jk} \cdot N_j^U , \quad \forall j,k \tag{19}$$

$$NZ1_{jk} \leq (1 - Z_{jk}) \cdot N_j^U , \quad \forall j,k \tag{20}$$

$$NZV_j = \sum_{k \in K} \sigma_k^2 \cdot NZ_{jk} , \quad \forall j \tag{21}$$

$$L_k \geq \sum_{j \in J} SZ_{jk} + \sum_{j \in J} t2_{jk} \cdot Z_{jk} - R_k , \quad \forall k \tag{22}$$

$$XZ_{ijk} \geq 0 , \quad SZ_{jk} \geq 0 , \quad SZ1_{jk} \geq 0 , \quad NZ_{jk} \geq 0 , \quad NZ1_{jk} \geq 0 , \quad NZV_j \geq 0 , \forall i,j,k \tag{23}$$

4. Solution Algorithm
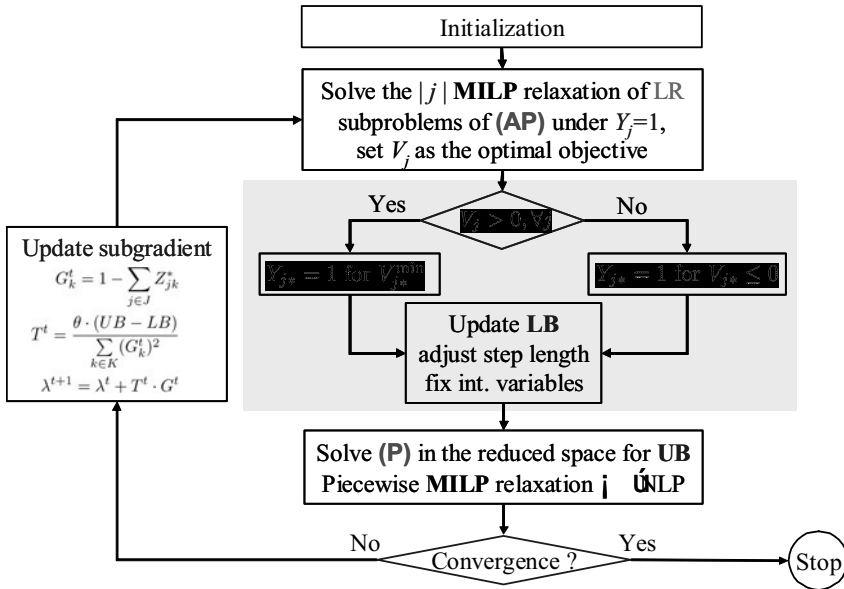
Figure 1. Lagrangean Relaxation Algorithm

To effectively solve the proposed MINLP model, a global optimization algorithm based on Lagrangean relaxation is developed. The algorithm flowchart is given in Figure 1. The basic idea of this algorithm is to consider the alternative formulation (AP) of the model. Next, we dualize the assignment constraint (10) to allow decomposing the entire problem based on DC j. To solve each subproblem, we used piece-wise linear approximation to underestimate the square root terms. We should note that the entire solution algorithm requires at least an MILP solver; the NLP solver is not required. Due to the duality gap, this algorithm stops after a finite number of iterations. As will be shown in the computational results, the dual gaps are quite small.

5. Results

To illustrate the application of the proposed model, we consider an industrial gas supply chain (liquid oxygen-LOX) with two plants, three potential DCs and six customers. The associated superstructure, as well as the optimal network structures with and without considering inventory cost, is given in Figure 2. The results show that although inventory cost only make up less than 20% of the total cost, it is necessary to take into account in the supply chain design. Comparison of the performance of the proposed

algorithms and commercial MINLP and global optimizers for medium and large scale instances (5 plants, 50 DCs, 150 markets) are given in Table 1. The advantage of using the Lagrangean relaxation algorithm for solving the proposed model can be clearly seen.
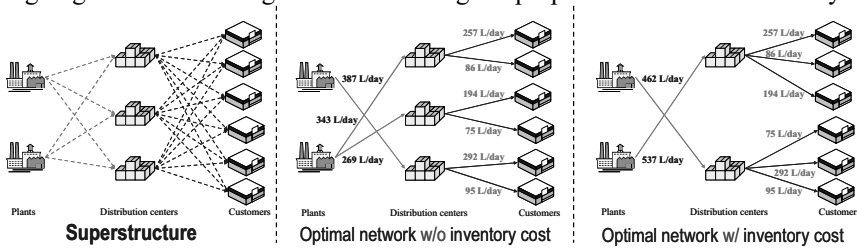
Figure 2. Optimal network structure for the LOX supply chain

Table 1. Comparison of the performance of the algorithms for medium and large scale instances

i	j	k	Solve (P0) directly with BARON				Solve (P2) with CPLEX for at most 1 hour, then solve (P1) with DICOPT or SBB				Lagrangean Relaxation Algorithm				
							DICOPT		SBB						
			Solution	LB	Gap	Time (s)	Solution	Time(s)	Solution	Time(s)	Solution	Global LB	Global Gap	Time (s)	Iter.
2	20	20	1,889,577	1,159,841	62.92%	36,000	1,820,174	140.3	1,813,541	163.6	1,776,969	1,775,957	0.06%	175.0	11
5	30	50	---*	---*	---*	36,000	---**	36,000	---**	36,000	4,417,353	4,403,582	0.31%	3,279	24
10	50	100	---*	---*	---*	36,000	---**	36,000	---**	36,000	7,512,609	7,477,584	0.47%	27,719	42
20	50	100	---*	---*	---*	36,000	---**	36,000	---**	36,000	5,620,045	5,576,126	0.79%	27,748	53
3	50	150	---*	---*	---*	36,000	---**	36,000	---**	36,000	12,291,296	12,276,483	0.12%	16,112	32

* No solution or bounds were returned due to solver failure.; ** No solution was returned after 10 hours

6. Conclusion

In this paper, we present an MINLP model that determines the optimal network structure, transportation and inventory levels of a multi-echelon supply chain with the presence of customer demand uncertainty. The well-known guaranteed service approach is used to model the multi-echelon inventory system. The risk pooling effect is also taken into account in the model by consolidating the demands in the downstream nodes to their upstream nodes. To solve the resulting MINLP problem effectively for large scale instances, a decomposition algorithm, based on Lagrangean relaxation and piecewise linear approximation was proposed. Computational experiments on large scale problems show that the proposed algorithm can obtain global or near-global optimal solutions (typically within 1% of the global optimum) in modest computational expense without the need of a global optimizer.

References

G. Eppen, 1979, Effects of centralization on expected costs in a multi-echelon newsboy problem. Management Science, 25, (5), 498.

M. L. Fisher, 1985, An application oriented guide to Lagrangian relaxation. Interfaces 1985, 15, (2), 2.

F. Glover, 1975, Improved Linear Integer Programming Formulations of Nonlinear Integer Problems. Management Science, 22, (4), 455.

S. C. Graves, S. P. Willems, 2005, Optimizing the supply chain configuration for new products. Management Science, 51, (8), 1165.

I. E. Grossmann, I. E., 2005, Enterprise-wide Optimization: A New Frontier in Process Systems Engineering. AIChE Journal, 51, 1846.

P. H. Zipkin, 2000, Foundations of Inventory Management. McGraw-Hill: Boston, MA, 2000.

Unit Slots Based Short-Term Scheduling for Multipurpose Batch Plants

Naresh Susarla, Jie Li, I. A. Karimi

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576

Abstract

In chemical process industries, scheduling of multipurpose batch plants is challenging and has received tremendous attention so far. Recently, Susarla et al. (2009) explored the concept of unit- (asynchronized) slots for scheduling multipurpose batch processes and successfully reduced number of slots required for the optimal solution of a given problem. In this paper, we extend the work of Susarla et al. (2009) and utilize unit-slots to formulate a continuous-time mixed integer linear programming (MILP) model for the short-term scheduling of multipurpose batch processes. In addition to Susarla et al. (2009), our model is capable of handling various utility resources (other than the processing units and material resources) and sequence dependent changeover/setup times. Also, our model is capable of handling various storage configurations explicitly (Classes: UIS, LIS, and FIS with policies: UW, LW, and NW, Liu & Karimi, 2007). We demonstrate the performance of our model through an extensive numerical evaluation with some of the best known models from the literature. This rigorous comparison further elucidates that our model uses fewer binary variables, continuous variables and constraints.

Keywords: Scheduling, multipurpose batch plants, slot based formulations, unit slots, MILP.

1. Introduction

Short term scheduling is a regular task in any batch-wise manufacturing facility. In particular, multipurpose batch plants (MBPs) produce a number of products with the limited and shared resources. Inherent complexity of handling shared resources induces both opportunities and challenges for the manufacturer. Attempts to address these challenges have started way back in 1980's. Mendez et al. (2006) present excellent review highlighting the current approaches and associated challenges.

The scheduling problem primarily revolves around the allocation of the equipments and other resources and then aligning various tasks over time. Apart from the raw materials, resources in MBPs generally include other utilities such as: steam, electricity, cooling water, manpower (operators), etc. However, very limited work is reported in the literature that considers utilities other than equipment and raw materials in their mathematical modeling. All the previous attempts show that different types of time representation chosen have a lasting impact on the MILP models. Three popular approaches for modeling time in the literature are slot based, event based, and sequence

based (precedence based). Recently, Susarla et al. (2009) presented a concise representation of the prevalent continuous time modeling techniques.

For slot based representation, the scheduling horizon is modeled in terms of sequential time blocks or time slots of unknown lengths. In literature, slot based approach appears in two different styles of implementation that are process slots (Sundaramoorthy & Karimi, 2005) and unit slots (Susarla et al., 2009). Events based approach defines random points over time to which different tasks are associated. Two different ways of implementing this approach, global event points (Maravelias & Grossmann, 2003; Castro et al., 2004) and unit-specific event points (Janak et al., 2004) are popular in the literature. Similar to the process slots, global event points define time points that are common across all the units. In unit-specific event points, the time points are not common and are defined independently on each unit. Li et al. (2009) show with the help of several examples the fallibility of unit-specific event based models in accounting for shared resources. The sequence-based approach mainly uses either direct or indirect sequencing of task-pairs on various units.

Susarla et al. (2009) proposed a novel continuous-time MILP model for the short-term scheduling of MBPs employing unit slots. They introduced a set of task sequencing constraints to know the relative timings of the tasks in order to handle sharing of resources and thus eliminate the need of an extra binary variable. However, they consider the only resources as the equipments and the raw materials. Also, they assumed the changeover/setup times involved (if any) to be lumped up with the batch processing times. In this work, we extend their model considering other utilities apart from equipment and raw materials. We also explicitly model the sequence dependent changeover/setup times over each unit. Consequently, this enhances the generality of the model for its application over a broader variety of problems. As we show later, even the introduction of more constraints our model preserves its superiority over other reported models in the literature.

2. Problem Statement

A multipurpose batch facility produces a number of products by performing I tasks ($i = 1, 2, …, I$) sharing a set of J batch processing units ($j = 1, 2, …, J$) that involves S material states ($s = 1, 2, …, S$) each with a dedicated storage s. Production in the plant follows a set of processes that is described by the given recipe diagram. In most of the cases, resources used at each step for processing the material also involve other utilities such as HP/LP steam, cooling water, etc. A generalized recipe diagram clearly depicts the sequence of processes, specific utilization of resources (materials, equipments, and utilities) at each step, and their utilization coefficients, UC (σ_{sij}). Each material state s in a multipurpose batch plant has a specified storage capacity (UIS/LIS/NIS) and wait policy (UW/LW/NW). The scheduling problem in MBPs can be described as follows. Given the (1) generalized recipe diagram (GRD), (2) processing units (J), their suitable tasks (I), and processing capacity limits (3) storage capacities and wait policies, (4) required utilities and their availabilities, (5) sequence-dependent changeover/setup times, and (6) market price of each material state, we need to determine (1) the optimal sequence of the tasks, (2) their batch sizes, and (3) inventory profiles. For this, we assume (1) deterministic scenario i.e., no disruptions whatsoever, (2) batch size dependent processing times, and (3) negligible unit to unit transfer times. The objective for this scheduling is mainly to either maximize the revenue for a given scheduling horizon [0, H] or minimize the makespan of the products to meet a given product demand.

3. MILP Formulation

We consider time to be continuous on each unit j and storage S and model it in terms of K ($k = 1, 2, \ldots, K$) contiguous slots of unknown and arbitrary lengths within the given horizon $[0, H]$. We represent the end time of slot k on processing unit j as T_{jk} [$k = 0, 1, 2, \ldots, K$; $T_{j0} \geq 0$; $T_{jK} \leq H$; $T_{jk} \geq T_{j(k-1)}$, $1 \leq k \leq K$]. Similarly, T_{sk} denotes the end time of slot k on storage S. The time before the slot 1 starts is slot 0 ($k = 0$). So, a slot k on a unit (processing, j, and storage, S) starts at time $T_{j(k-1)}$, ends at time T_{jk}, and has a length [$T_{jk} - T_{j(k-1)}$]. Also, we assume that the real task begins at the start and may end at any time during the slot. Thus, the idle time (if any) is always towards the end of a slot. Moreover, a unit performs at most one task in each slot. However, it is quite possible that a unit performs no task during a slot or remains idle. To model this idling of units we define a zero (or an idle) task, $i = 0$. To schedule various tasks on different units, we now define one binary variable and two 0-1 continuous variables that denote respectively the start, end and continuations of a task on a unit slot as follows,

$$ys_{ijk} = \begin{cases} 1 & \text{if unit } j \text{ begins a new batch of task } i \text{ at } T_{jk} \\ 0 & \text{Otherwise} \end{cases} \qquad 1 \leq j \leq J, i \in \mathbf{I}_j, 0 \leq k < K$$

$$ye_{ijk} = \begin{cases} 1 & \text{if unit } j \text{ ends a batch of task } i \text{ within slot } k \\ 0 & \text{Otherwise} \end{cases} \qquad 1 \leq j \leq J, i \in \mathbf{I}_j, 1 \leq k \leq K$$

$$y_{ijk} = \begin{cases} 1 & \text{if unit } j \text{ continues a batch of task } i \text{ at } T_{jk} \\ 0 & \text{Otherwise} \end{cases} \qquad 1 \leq j \leq J, i \in \mathbf{I}_j, 0 \leq k < K$$

y_{ijK} is undefined, y_{ij0} is known and fixed. If a batch of task $i > 0$ is unfinished at time zero, and must continue, then $y_{ij0} = 1$, otherwise $y_{ij0} = 0$. This also sheds light on the fact that we do not force that each batch must start and end during a single slot, but it can span several slots.

Using the above, Susarla et al. (2009) developed several constraints for task & batch allocation, batch sizing, time sequencing, and material balancing. Some of these constraints are given below. For the batch to unit allocation, the following unit status balance equation is used

$$ye_{ijk} = [y_{ij(k-1)} + ys_{ij(k-1)}] - y_{ijk} \qquad\qquad 1 \leq j \leq J, 1 \leq k < K \qquad (1)$$

Similarly, for determining batch sizes for each batch the following balance on the amount of task i over slot k give

$$BO_{ijk} = b_{ij(k-1)} + [B_{ij}^L ys_{ij(k-1)} + \Delta BI_{ij(k-1)}] - b_{ijk} \qquad 1 \leq j \leq J, i \in \mathbf{I}_j, i > 0, 1 \leq k \leq K \qquad (2)$$

where, BO_{ijk} is amount of material that is released at the end of the ongoing task i during slot k, ΔBI_{ijk} is differential amount of material that is consumed at the beginning of slot k by task i, and b_{ijk}. Again, a balance on the remaining batch processing time (t_{jk}) yields the following equation

$$t_{j(k+1)} \geq t_{jk} + \sum_{i \in \mathbf{I}_j} [(\alpha_{ij} + \beta_{ij} B_{ij}^L) ys_{ijk} + \beta_{ij} \Delta BI_{ijk}] - (T_{j(k+1)} - T_{jk}) \qquad 1 \leq j \leq J, 0 \leq k < K \quad (3)$$

In addition to the aforementioned constraints, the following sequencing constraints are required to ensure proper mass balance

$$T_{sk} \leq T_{jk} + H[1 - \sum_{i \in \mathbf{I}_j, \sigma_{sij} < 0} ys_{ijk}] \quad 1 \leq j \leq J, \quad \sum_{i \in \mathbf{I}_j, \sigma_{sij} < 0} \sigma_{sij} < 0, \ 0 \leq k < K, s \ni I_s^U \text{ is LIS} \quad (4a)$$

$$T_{sk} \geq T_{jk} - H[1 - \sum_{i \in \mathbf{I}_j, \sigma_{sij} > 0} ye_{ijk}] \quad 1 \leq j \leq J, \quad \sum_{i \in \mathbf{I}_j, \sigma_{sij} > 0} \sigma_{sij} > 0, \ 1 \leq k < K, s \ni I_s^U \text{ is LIS} \quad (4b)$$

The above constraints, along with few more complete the model proposed by Susarla et al. (2009), but is incapable of handling utilities/resources other than raw materials and the processing equipments. However, chemical process industries in general and multipurpose batch plants in particular, require several other utilities/resources like- HP / LP steam for heating purposes, cooling water / ethylene glycol for cooling purposes, etc. While monitoring the consumption and a balance on these utilities are necessary, they are not addressed by Susarla et al. (2009). Therefore, in this paper, we extend that formulation by incorporating constraints for utility balance. For this, we first define $T_{u,k}$ as the end time of the slot on the utility usage, u. Also, we define $U_{u,k}$ as the rate of utility u in use during slot k. Now, as each utility is associated with certain processing task, the timings for the slots on these utilities have to be same as that of the processing units. So,

$$T_{uk} = T_{jk} \hspace{5cm} 1 \leq k \leq K \hspace{3cm} (5)$$

After associating the timings of utility usage, we now ensure the resource usage balance for each utility u. We model the utility usage based on its rate of supply. As we allow an unfinished batch to continue at time zero, the utility usage at the beginning of the horizon (U_{u0}^{in}) is known for every unit j. The rate of utility usage is determined by the size of a batch. The following equations make a balance on the rate of utility available at the end of any slot k.

$$U_{u1} = U_{u0}^{in} + \sum_{i \in \mathbf{I}_r, i > 0} \sum_{j \in \mathbf{I}_j} (\gamma_{ij} + \delta_{ij} B_{ij}^L) ys_{ij1} + \delta_{ij} \Delta BI_{ij1} \hspace{3cm} (6a)$$

$$U_{uk} = U_{u(k-1)} - \sum_{i \in \mathbf{I}_r, i > 0} \sum_{j \in \mathbf{I}_j} (\gamma_{ij} ye_{ijk} + \delta_{ij} BO_{ijk}) + \sum_{i \in \mathbf{I}_r, i > 0} \sum_{j \in \mathbf{I}_j} ((\gamma_{ij} + \delta_{ij} B_{ij}^L) ys_{ijk} + \delta_{ij} \Delta BI_{ijk})$$

$$1 < k < K \hspace{3cm} (6b)$$

$$U_{uK} = U_{u(K-1)} - \sum_{i \in \mathbf{I}_r, i > 0} \sum_{j \in \mathbf{I}_j} \gamma_{ij} ye_{ijK} + \delta_{ij} BO_{ijK} \hspace{3cm} (6c)$$

The formulation so far assumes the changeover/setup times to be lumped up with batch processing times. However, with a small extension this formulation can handle sequence dependent changeover/setup times.

We define a maximum idling parameter, Φ, for each unit j that denotes the maximum number of consecutive slots for which a unit can remain idle. We start with $\Phi = 1$ and increase its value by 1 until there is no change in the objective function. Now, a sequence dependent changeover/setup time ($\tau_{i'ij}$) is to be incorporated between two consecutive real tasks. For this, we demand the start time of a real task to be greater than or equal to the sum of processing time of the immediately preceding real task and the associated changeover/setup time.

$$T_{jk} \geq T_{jk'} + (\alpha_{i'j} + \beta_{i'j} B_{i'j}^L) ys_{i'jk'} + \beta_{i'j} \Delta BI_{i'jk'} + \tau_{i'ij} ys_{ijk} - H(1 - ys_{i'jk'}) - H \sum_{\substack{i'' \in \mathbf{I}_j, i'' > 0 \\ i'' \neq i, i'' \neq i'}} \sum_{k' < k'' < k} ys_{i''jk''}$$

$$i, i' > 0, i \neq i', i, i' \in \mathbf{I}_j, \ 1 < k < K, \ k > k', \ (k - k') \leq \Phi, \ \tau_{i'ij} > 0 \qquad (7)$$

Note that eq. 7 imposes the required changeover/setup time ($\tau_{i'ij}$) between two consecutive tasks and is relaxed otherwise.

Now, considering the suitable wait policy (UW/LW/NW) for each material state s, we implement the following constraints.

$$t_{j(k+1)} \leq t_{jk} + \sum_{i \in \mathbf{I}_j}[(\alpha_{ij} + \beta_{ij}B_{ij}^L)ys_{ijk} + \beta_{ij}\Delta BI_{ijk}] - (T_{j(k+1)} - T_{jk}) + H[1 - \sum_{i \in \mathbf{I}_j, \sigma_{sij} > 0} yr_{ij(k+1)}]$$

$$1 \leq j \leq J, \ 0 \leq k < K \qquad (8a)$$

$$T_{sk} \leq T_{j(k-1)} + t_{j(k-1)} + \sum_{i \in \mathbf{I}_j}[(\alpha_{ij} + \beta_{ij}B_{ij}^L)ys_{ij(k-1)} + \beta_{ij}\Delta BI_{ij(k-1)}] + \sum_{i \in \mathbf{I}_j, \sigma_{sij} > 0}[w_{ij}ye_{ijk}] + H[1 - \sum_{i \in \mathbf{I}_j, \sigma_{sij} > 0} ye_{ijk}]$$

$$1 \leq j \leq J, \ \sum_{i \in \mathbf{I}_j, \sigma_{sij} > 0} \sigma_{sij} > 0 \ , \ 1 \leq k < K, \ s \ni I_s^U \text{ is LIS} \qquad (8b)$$

With the aforementioned equations in association with the model from Susarla et al. (2009) completes our model for the problems with utility constraints and sequence dependent changeover times.

4. Model Evaluation and Results

We present two case studies from Maravelias & Grossmann (2003), to evaluate the performance of our model. Fair and unbiased evaluations of different models require (Karimi et al., 2004; Susarla et al., 2008) a careful attention to the factors such as hardware, software, and the operating system. For our evaluation, we used CPLEX 11/GAMS 22.8 on a Dell precision PWS690 workstation with Intel® Xeon® 3 GHz CPU, 16 GB RAM, running windows XP Professional x64 Edition. Also, we solve various scenarios of each problem, demonstrating the robustness of our model.

4.1. Case study-1

Maravelias & Grossmann (2003) first used this example to illustrate the handling of utility constraints. The relevant data for this example can be obtained from the original paper. It involves 3 units, 4 tasks, 7 material states, and 2 utilities (cooling water and high pressure steam). We solve this example for different scenarios based on the rate of availability of utilities. In scenario A, we assume that both the utilities are available at a rate of 40Kg/min and scenario B we assume it to be 30 Kg/min. Also, we solve this example for two scheduling objectives: maximizing revenue and minimizing makespan. Table 1 consolidates the model statistics for both (ours and M&G) models. Clearly, our model takes fewer binary variables, continuous variables, constraints, and non-zeros in all the scenarios. This explains the simplicity of our model.

4.2. Case study-2

This example is from Maravelias & Grossmann (2003) (Example 3). While, the relevant data can be found in the original paper, it involves 6 units, 10 tasks, 14 material states, and 3 utilities. This example demonstrates the capability of these models in handling zero wait policy for some of the intermediate material states and various storage capacities for others. Again, we consider two different scheduling horizons to solve two different problems. Our model performs better and again requires fewer binaries, continuous variables, constraints, and non- zeros.

Table 1 Model and solution statistics for Case studies 1 & 2

Model	K	CPU (s)	Nodes	Objective	Binary Var	Continuos Var	Constraints	Nonzeros
Case Study 1: Scenario A - Revenue Maximisation								
Ours	7	1.72	1416	$5,904	54	365	452	1676
M & G	7	1.70	975	$5,904	84	484	1052	4117
Case Study 1: Scenario B - Revenue Maximisation								
Ours	6	0.26	126	$5,227.78	45	310	378	1396
M & G	6	0.15	67	$5,227.78	72	415	901	3356
Case Study 1: Scenario A - Makespan Minimisation								
Ours	8	6.67	6053	8.5 h	63	452	570	2057
M & G	8	4.30	1463	8.5 h	96	610	1212	5073
Case Study 1: Scenario B - Makespan Minimisation								
Ours	7	0.53	405	9.025 h	54	393	491	1765
M & G	7	0.69	292	9.025 h	84	534	1061	4241
Case Study 2: Scenario A (H = 12h) - Revenue Maximisation								
Ours	9	20.30	8037	$13,000	128	883	1366	4667
M & G	9	14.80	5315	$13,000	180	1009	2252	8033
Case Study 2: Scenario B (H = 14h) - Revenue Maximisation								
Ours	10	143.00	41933	$16,350	144	985	1535	5251
M & G	10	262.03	31906	$16,350	200	1121	2503	9240
M & G = Maravelias & Grossmann (2003)								

5. Conclusions

This paper successfully utilizes the concept of unit slots, originally proposed by Susarla et al., (2009) and extends their formulation to consider utility handling constraints, sequence dependent changeover/setup times. While additional constraints are added for the said extension, the model preserves its superiority among the equals. Another critical point in this paper is that it does not add any extra binary variable to know the relative positions of the tasks as commonly done in slot based formulations.

References

Susarla, N., Li, J., Karimi, I. A., A Novel Approach for Scheduling Multipurpose Batch Plants using Unit Slots, *Submitted to AIChE Journal*, 2009.

Liu, Y., Karimi, I. A., Scheduling multistage, multiproduct batch plants with nonidentical parallel units and unlimited intermediate storage, *Chemical Engineering Science*, 2007, 62, 1549-1566.

Mendez, C.A., J. Cerdá, I. E. Grossmann, I. Harjunkoski, and M. Fahl, State-Of-The-Art Review of Optimization Methods for Short-Term Scheduling of Batch Processes, *Computers & Chemical Engineering*, 2006, 30, 913-946.

Sundaramoorthy, A., Karimi, I. A., A simpler better slot-based continuous-time formulation for short-term scheduling in multipurpose batch plants, *Chemical Engineering Science*, 2005, 60, 2679-2702.

Maravelias, C. T., Grossmann, I. E., New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants, *Industrial & Engineering Chemistry Research*, 2003, 42, 3056-3074.

Janak, S. L., Lin, X., Floudas, C. A., Enhanced continuous-time unit-specific event-based formulation for short-term scheduling of multipurpose batch processes: Resource constraints and mixed storage policies, *Industrial & Engineering Chemistry Research*, 2004, 43, 2516-2533.

Li, J., Susarla, N., Karimi, I. A., The Fallibility of Some Unit-Specific Event-based Models for The Short-term Scheduling of Noncontinuous processes, *Submitted to Industrial & Engineering Chemistry Research*, 2009.

Castro, P. M.; Barbosa-Povoa, A. P.; Matos, H. A.; Novais, A. Q., Simple Continuous-Time Formulation for Short-Term Scheduling of Batch and Continuous Processes. *Industrial & Engineering Chemistry Research* 2004, *43* (1), 105-118.

1995

Linking Marketing and Supply Chain Models for Improved Business Strategic Decision Support

José Miguel Laínez,[a] Gintaras V. Reklaitis,[b] Luis Puigjaner[a]

[a]*Department of Chemical Engineering, Universitat Politècnica de Catalunya, Barcelona, Spain, Luis.Puigjaner@upc.edu*
[b]*School of Chemical Engineering, Purdue University, West Lafayette, USA*

Abstract

Nowadays, a Supply Chain (SC) management model incorporating business strategic components is becoming of paramount importance to gain a competitive edge in the market place. To be successful, the enterprise model has to contemplate not only the supply chain, but also the demand chain. Understanding the market and customer behavior is extremely crucial for developing a good business policy. Marketing is a boundary-spanning activity, linking selling entities with buyers and intermediate channels. To operate most effectively, marketing activities must be coordinated with other functional areas of the firm. The marketing - SC management interface is an issue that deserves further research. Business managers should evaluate the existing trade-off between marketing and SC planning decisions in order to enhance the performance of the overall business metric: the shareholders value. Recently, there is a significant trend driving business managers to implement marketing science models for reaching more rational and holistic decisions. In this work, a mathematical model is presented for the enterprise that accounts for the three main business functionalities (i.e. operations, finances and marketing). The posed problem is tackled by developing a holistic MINLP model which optimizes in tandem the SC and marketing strategic decisions. Moreover, a financial model that allows to quantitatively assessing the enterprise value is also incorporated. Finally, main advantages of coordinated decisions are discussed through an illustrative example.
Keywords: enterprise wide optimization, marketing models, business decision support systems

1. Introduction

The business strategy is most often modeled as a hierarchical process in which functional strategies, such as operations, logistics, marketing, and finance are driven by a higher level strategy (Fine and Hax, 1985). A key element of the strategic framework involves coordinating functional-level plans to work in concert so as to achieve the overall business strategy rather than to locally optimize outcomes for individual functions, business units, plants, or stores. One of the primary challenges in implementing an effective strategy involves achieving consensus within the business organization. Unfortunately, while this concept is clearly sound on a conceptual level, actual implementation is typically very difficult (Bozarth and Berry, 1997).

One important strategic issue that needs more research efforts is the integration of supply chain (SC) production-distribution operations and marketing activities. Although several authors have highlighted the conflicting goals of supply chain (SC) and marketing managers (Eliashberg and Lilien, 1993; Shapiro, 2007), it is still typically

assumed that under a decentralized decision making scheme, marketing decisions are made first; determining demand forecasts which are later considered by the SC model to support production-distribution related decisions. By deploying this sequential procedure, the firm may be significantly under-estimating its overall performance. Integrating marketing and operations is a challenge in any business, since there is a natural tension between these two functional areas (Bozarth and Berry, 1997). At best, the tension between these two functions results in a dampening of marketing's tendency to over-promise to lure customers and a push on operations to move beyond an internal focus on reducing costs without a clear vision of end-consumer needs.

Usually, the primary objective of marketing function is maximizing revenues creation by satisfying customers through the products and services offered. On the other hand, SCM's focus is on the synchronization of production and distribution activities along the different entities comprising the SC network. The main objective is typically to minimize the total SC cost. In general, conflicts arise between marketing and SC because of these contrasting performance indicators which eventually are used to develop incentive structures for managers and their corresponding employees. For instance, one classical conflict between these two functions is the one associated to the inventory management. SC managers strive to keep low stock levels, while marketing managers long for high stock levels to guarantee that most of customer orders are met, thus improving revenue generation. Nevertheless, the enterprise main goal is to create and maximize shareholders value which actually is a function of revenues, cost and other economic factors. Consequently, business managers are in need of an integrated analytical decision support tool that is capable of appraising the trade-off between operations and marketing while evaluating and maximizing shareholders value.

Nowadays, there are more and more companies that are continuously searching for competitive advantages in order to get a better position in markets. One way of doing so may be by aligning functionalities strategic/tactical decisions towards the optimization of the overall business performance metric. In this work, we present a novel approach to address this challenge. We develop a MINLP model that tackles SC network design and marketing strategic decisions in tandem. Then, such model is coupled with a capital budgeting formulation which allows calculating shareholders value by means of the discounted-free-cash-flow (DFCF) method.

2. The marketing strategic model

Today firms have access to more market and customer data than they can use. Having too much data without the models and systems for analyzing what is important and what can be discarded can be as bad as or even worse than having too little data. Increasingly marketing managers are being asked to clear the same budget-justification hurdles imposed on other types of investment firms make. It is not surprising, therefore, that more managers are seeking help in turning their data and knowledge into improved decision making (Lilien and Rangaswamy, 2002). The systematic translation of data and knowledge into a normative and/or descriptive model that can be used for marketing decision support is what it is recently regarded as marketing engineering.

Here, BRANDAID, a marketing engineering contribution, is used. This flexible strategic marketing mix model is basically comprised of three sub-models, namely: advertising sub-model, pricing sub-model and sales force sub-model. It is noteworthy

that the model is (i) modular, so managers determine which sub-models to take into account according to their needs and (ii) customizable, since managers may introduce into the sub-models those functions that best describe their own marketing processes. The model is briefly presented next following the work of Little (1975). Two sub-models are considered in this paper; however other marketing influences/activities can be handled in the same manner.

2.1. The main model structure.

The model is based on the two following expressions. Eq. (1) expresses firm's target market demand (dem_{it}) as a function of market share (ms_{it}) and total market demand (D_t), while Eq. (2) computes market share as a reference value (S_o) modified by the effect of marketing activities (pricing, advertising, promotions, sales force) and other considered influences (e_{it}) such as competitors actions.

$$dem_{it} = ms_{it}D_{it} \qquad \forall i,t \tag{1}$$

$$ms_{it} = S_o \prod_{ma} e_{i,ma,t} \qquad \forall i,t \tag{2}$$

2.2. Advertising sub-model

This module assumes that there exists an advertising rate that maintains sales at target level. When the advertising rate is above reference, sales are assumed to increase; otherwise they decrease. The next equations describe this procedure. We assumed *ma=1* to describe advertising activities.

$$e_{i1t} = \varpi_1 e_{i1t-1} + (1 - \varpi_1)A_{it} \qquad \forall i,t \tag{3}$$

$$A_{it} = f(h_t, k_t, AdvExp_t) \qquad \forall i,t \tag{4}$$

$$A_{it} = \frac{h_{it}k_{it}AdvExp_{it}}{h_{io}k_{io}AdvExp_{io}} \qquad \forall i,t \tag{5}$$

Here, ϖ_1 represents the carryover effect of advertising per period. The advertising rate (A_{it}) is the rate of messages delivered to customers by exposure in media. This rate is usually considered as a function of media efficiency (h_{it}), copy-effectiveness (k_{it}) and advertising expenditures ($AdvExp_{it}$). As previously mentioned, this function can be estimated by each particular case by examining the data available regarding this marketing activity. Hence, particular and current market-media trends can be taken into account by constructing an adequate descriptive model. For our illustrative example a function similar to Eq. (5) is used. In that equation, h_{io}, k_{io}, and $AdvExp_{io}$ are reference values of the above-mentioned parameters.

2.3. Pricing sub-model

Price elasticity of demand is considered by this sub-model. Again, more complicated descriptive models can be used to describe the specific product behavior of price-demand. We assumed *ma=2* to describe pricing activities. Here the following linear function is used. ε_t is the elasticity coefficient.

$$price_o(1-e_{i2t}) = \varepsilon_t(price_{it} - price_o) \quad \forall i,t \tag{6}$$

3. The SC design – planning model.

The design-planning approach is based on the work developed by Laínez et al. (2007). In this formulation, a four echelon SC is considered as shown in Figure 1. In this work, the equation that expresses that part of the demand can be left unsatisfied because of limited production capacity becomes one of the integrating equations as stated in section 5.
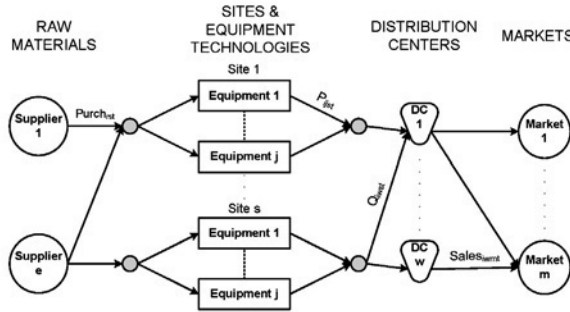
Figure 1: Supply chain model structure

4. The financial formulation

The financial component of the problem is tackled through the inclusion of a set of constraints that characterize economic issues, such as, payments to providers, loans, pledging decisions, etc. Also liquidity is controlled by this model. Furthermore, the Corporate Value (CV), is the objective function used. The CV of a company is a function of four factors: (i) investment, (ii) cash flows, (iii) economic life, and (iv) capital cost. Specifically, our work applies the DFCF method to compute the CV. Such method rates an entire company by determining the present value of its future free cash flows and discounting them, taking into account the appropriate capital cost during the evaluating time horizon. Capital cost is calculated using the weighted average method (WACC) which considers the firm's overall equity and debt.

Free cash flow at every period t (FCF_t) is defined by a function that depends on net operating profit after taxes, change in net working capital (ΔNWC_t), net change in investments ($NetInvest_t$) and advertising expenditures ($AdvExp_t$). Notice that Eq. (7) is already integrating marketing with finances through advertising expenditures. Eq. (8) is to compute the CV. For financial formulation details the reader is referred to Laínez et al. (2007)

$$FCF_t = Profit_t(1-trate) - (\Delta NWC_t + NetInvest_t + AdvExp_t) \quad \forall t \tag{7}$$

$$CV = \sum_{t=0}^{T} \frac{FCF_t}{(1+WACC_t)^t} - NetDebt_0 \tag{8}$$

5. Integration among models

Additionally to Eq. (7), integration among models is carried out by demand and revenue ($ESales_t$). As it can be observed a better balance between the demand generated by means of advertising and the available capacity proposed by the SC model can be achieved by utilizing Eq. (9).

$$\sum_w Sales_{iwt} \leq dem_{it} \qquad \forall i,t \tag{9}$$

$$ESales_t = \sum_i \sum_w Sales_{iwt} \, price_{it} \qquad \forall t \tag{10}$$

Notice that price and demand are problem variables instead of parameters and products i should be transported to final market from distribution centers w.

6. Illustrative example

The advantages of the proposed approach are demonstrated by solving an illustrative SC design-planning problem; which contains three processing sites (*S1-S3*), three distribution centers (*W1-W3*) and three market locations (*M1-M3*). A set of potential equipment technologies are assumed to be available for the processing sites. Three product families (*P1-P3*) can be manufactured on three different equipments types (*TA to TC*). A time horizon of 5 annual periods is considered. For all markets and products, the elasticity coefficient and the reference advertising expenditures are assumed to be equal to 0.85 and 3×10^6 c.u. respectively. It takes 4 CPU seconds to reach the optimal solution for the integrated problem on an Intel Core 2 Duo at 2.0 GHz computer using DICOPT.

For comparison purposes, the problem has been also solved using the traditional sequential approach (SA) described in section 1. First marketing decisions are taken and then the SC and financial decisions are determined. The marketing decisions are typically obtained by optimizing net revenues (subtracting advertising expenditures from total revenues), while the financial and SC decisions are computed by maximizing the CV. Numerical results show that the solution calculated by the integrated approach (IA) offers improved performance over SA. Certainly, the optimal expected CV from IA is 47% higher than the one obtained by utilizing SA. The IA obtains such performance by a 2% decrease in net revenues (Table 1). The two approaches also yield different SC design decisions: SA proposes to install all equipment technologies in sites *S1* and *S3,* and *TA* in *S2*, while the IA proposes to install merely *TA* and *TC* in location *S1* and all technologies in site *S3*. Figure 2 shows advertising expenditures and the product demand induced by marketing activities for both approaches.

Table 1. Performance comparison between approaches

Approach	Revenues	Advertising Expenditures	Net Revenues	Corporate value	Investment in capacity
Sequential	522.32×10^6	86.11×10^6	436.21×10^6	84.35×10^6	79.77×10^6
Integrated	512.94×10^6	85.15×10^6	427.79×10^6	123.93×10^6	43.42×10^6

*All figures in c.u.
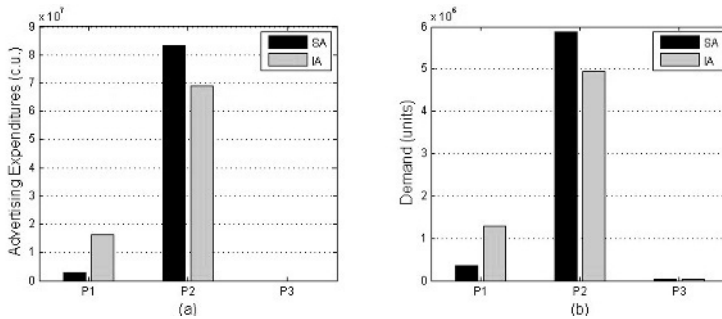
Figure 2. (a) Advertising expenditures and (b) Induced demand for both approaches

7. Conclusions

The intent of this paper is to motivate and draw attention to the need of further research in this kind of decision problems which are interfacing SC operations and marketing functions. A MINLP model integrating SC design/retrofitting, capital budgeting and marketing decisions is presented. Performance comparison with the traditional sequential decision approach is made, which demonstrates the significant economic benefits that the holistic approach provide. Our aim is to emphasize the relevance of a correct appraisal of the trade-off existing between the demand (which can be induced by marketing expenditures) and the SC capacity investments required to meet such demand. Finally, we would like to point out the great potential of response surface and data mining techniques in this field. Further work is focused on applying decomposition techniques so that (i) medium industrial size cases may be tackled using the proposed model and (ii) more complicated marketing models explored may be computed in acceptable time.

8. Acknowledgements

References

Bozart, C. C. and W. L. Berry, 1997. *Decision Sciences*, 28 (1) 121.

Eliashberg, J. and R. Steinberg, 1993. Marketing-Production Joint Decision-Making. In *Handbooks in Operation Research and Management Science*, Vol. 5, Marketing. Elsevier Science Publishers B.V.

Grant, J. L., 2003. *Foundations of Economic Value Added*; John Wiley and Sons

Laínez J.M., G. Guillén-González, M. Badell, A. Espuña and L. Puigjaner, 2007. *Ind. Eng. Chem Res.,* 46 (23), 7739.

Lilien, G. and A. Rangaswamy, 2001. *Interfaces*, 31(3) S1

Little, J.D., 1975. *Operations Research*, 23: 628.

Shankar, V., 2008. Strategic Marketing Decision Models for the Pharmaceutical Industry. In: *Handbook of Marketing Decision Models*. Springer: New York.

Shapiro, J. F., 2007, *Modeling the Supply Chain*. Duxbury.

Operation of the Argentinean Interconnected Electricity Network

Pablo Martínez, Ana María Eliceche

Departamento de Ingeniería Química, Universidad Nacional del Sur, PLAPIQUI-CONICET, Camino La Carrindanga km 7, 8000 Bahía Blanca, Argentina.

Abstract

A methodology to select optimally the electricity generation plants of the Argentinian interconnected network is developed. The electricity generation grid has different fossil fuel, hydroelectric and nuclear power stations. The power plants configuration connected to the grid and their operating loads are selected minimizing life cycle greenhouse gas emissions and operating cost simultaneously. A life cycle approach to estimate greenhouse gas emissions of thermoelectric, hydroelectric and nuclear power plants is followed. Binary operating variables represent discrete decisions to select which power plant is connected to the grid and the type of fossil fuel used. Continuous operating variables are introduced to select the optimal load for each power plant. A mixed integer linear programming problem is formulated and solved in GAMS. Significant reductions in green house emissions and operating cost are achieved simultaneously in the operation of the electricity network. Thus, a useful tool to support a decision-making process in the operation of a key energy sector is presented.

Keywords: Interconnected electricity network, operation, greenhouse emissions, cost.

1. Introduction

Environmental concerns have reached a high societal interest in a few years. Special interest has received greenhouse emissions of energy generation, claiming for environmental responsible energy policies. The main source of greenhouse emissions is the combustion of fossil fuels although greenhouse emissions are also present in the entire life cycle of many products or services, and electricity is not an exception. The upstream processes include raw material extraction, processing and distribution consuming energy, which have associated greenhouse emissions due to both fossil fuel consumption and fugitive emissions. Weisser (2007) presented an exhaustive work on life cycle greenhouse emissions in energy generation paying special attention to fossil fuel, nuclear and renewable energy technologies in the European Union and Japan. Hashim et al. (2005) studied the Ontario energy system minimizing CO_2 emissions. However, the authors have not considered the life cycle CO_2 emissions occurred in the upstream processes of each electricity generation option. The international price of CO_2 emissions is included in the objective function of the mixed integer linear programming problem formulated.

In the present work, life cycle green house emissions are estimated in the electricity generation and the limits of each power plant are extended, from raw material extraction to waste disposal. The Argentinean electricity grid has coal, fuel oil, gas oil and natural

gas driven thermoelectric plants, nuclear and hydroelectric plants. The Objective function minimized is a combined function that includes the operating cost and the benefits in the emission trade market for reducing the greenhouse life cycle emissions. The greenhouse gases international price was used to translate an environmental burden into an economic value. The network operating cost includes the costs of fuels and maintenance of each plant. Eliceche and Martínez (2007) successfully used a similar approach to select the operating conditions of a steam and power plant minimizing life cycle greenhouse emission of imported electricity and natural gas feedstock.

The methodology presented for the selection of the operating conditions of the Argentinian electricity network leads to significant reductions in operating cost and green house emissions simultaneously.

2. Electricity Network Modeling

2.1. Electricity Generation

The modeling of the interconnected system includes continuous and binary variables. The generation of each power plant is modeled as a fraction of its installed capacity. Binary variables represent discrete decisions and are introduced to select the type of fossil fuel used in a given thermoelectric plant, and to select which power plant (fossil-fuel based or not) is connected to the grid for a fixed yearly demand. The mix of all the electricity generated is injected into the electricity grid. The model considers the electricity generated by a certain power plant as a fraction of its maximum installed capacity, $\mathbf{G^{Max}}$:

$$\mathbf{G(q,f)} = \mathbf{G^{Max}(q,f)} \times \mathbf{d(q,f)} \qquad \forall \mathbf{q} \in \mathbf{F} \tag{1}$$

$$\mathbf{G(q)} = \mathbf{G^{Max}(q)} \times \mathbf{d(q)} \qquad \qquad \forall \mathbf{q} \in \mathbf{NF} \tag{2}$$

Where $\mathbf{G(q,f)}$ is the electricity generated, in *Gwh/yr*, by power plant $\mathbf{q}$ burning fossil fuel $\mathbf{f}$. All the fossil fuel driven power plants are included in the group $\mathbf{F}$. $\mathbf{G(q)}$ is the electricity generated by the power plant $\mathbf{q}$ which does not consume fossil fuels and belong to the group $\mathbf{NF}$ where nuclear and hydroelectric power plants are included. The variables $\mathbf{d(q,f)}$ or $\mathbf{d(q)}$ are the availability factor of each power plant. It express the ratio between the energy produced by a power plant in a certain period of time and the energy that it would be generated by the power plant working at its maximum installed capacity during the same period of time.

In order to select the fossil fuel used in a certain fossil fuel power plant, it is necessary to include binary variables in the mathematical model. The binary variable $\mathbf{y_{q,f}}$ takes the value 1 if the power plant $\mathbf{q}$ is burning the fossil fuel *f* and it is equal to 0 otherwise. The fact that a certain thermoelectric power plant could only work with an alternative fossil fuel, in a given time period, is modeled with the following equation:

$$\sum_{\mathbf{f}} \mathbf{y_{q,f}} \leq 1 \tag{3}$$

Binary variables $\mathbf{y_q}$ are defined for the group of non-fossil fuel plants NF, to select which hydroelectric or nuclear power plants are on or off during the operation. The electricity generated for any power plant could not be greater than its installed capacity and it cannot be lesser than a certain value imposed by the interconnected system:

$$G(q,f) \leq G^{Max}(q,f) \times \sum_f y_{q,f} \qquad \forall q \in F \tag{4}$$

$$G(q) \leq G^{Max}(q) \times y_q \qquad \forall q \in NF \tag{5}$$

$$d(q,f) \geq d_{q,f}^{LB} \times y_{q,f} \qquad \forall q \in F \tag{6}$$

$$d(q) \geq d_q^{LB} \times y_q \qquad \forall q \in NF \tag{7}$$

Equations 4 and 5 represent upper bounds on energy production from each plant **q**. The Eq. 4 ensures that electricity generation from power plant **q** is zero when no fossil fuel is assigned to the power plant and the plant is shut down. The Eq. 5 indicates that electricity production in non fossil fuel (**NF**) plant **q** is smaller or equal to its maximum capacity. Equations 6 and 7 set up the lower limits in the availability factors of each group of power plants. These lower limits establish the minimum quantity of electricity generated by a certain power plant **q**.

An upper limit on the availability factor is set up in the Equations 8 and 9.

$$d(q,f) \leq (1+\beta_q) \times d^a(q,f) \qquad \forall q \in F \tag{8}$$

$$d(q) \leq (1+\beta_q) \times d^a(q) \qquad \forall q \in NF \tag{9}$$

The superscript "**a**" indicates the current value of the availability factor for each power plant. The parameter β_q is the maximum increment allowed for the availability factor for each power plant in the time period considered.

A demand satisfaction constraint is shown in Eq. 10, where **D** is the entire network electricity demand for the time period considered:

$$\sum_{q \in F} \sum_f G(q,f) + \sum_{q \in NF} G(q) \geq D \tag{10}$$

The operating cost equation for the entire network follows, where $C_F(q,f)$ and $C_{NF}(q)$ are the operating cost of each power plant in *US\$/Gwh*, including fuels and maintenance costs.

$$C_G = \sum_{q \in F} \sum_f G(q,f) \times C_F(q,f) + \sum_{q \in NF} G(q) \times C_{NF}(q) \tag{11}$$

2.2. Greenhouse Emissions Quantification

A life cycle approach estimates greenhouse gas emissions of each thermoelectric, hydroelectric and nuclear power plant. The life cycle approach considers emissions during the life cycle of each electricity generation plant, from raw material extraction to waste disposal including the generation step itself. The quantification of greenhouse gases emissions (**GHG**) is estimated using emission factors, which express the mass of a certain greenhouse gas **k** emitted by unit of electricity generated. Greenhouse gases include CO_2, N_2O, CH_4, SF_6 and CFCs, each of them having different heat-trapping properties. To compare their effects on the atmosphere the Global Warming Potential, the **gwp** factors are used. Global Worming Potential express the ability of a greenhouse

gas to trap heat in the atmosphere relative to an equal amount of carbon dioxide, thus **gwp** factor is expressed in *mass of CO_2 equivalent/mass of GHG* k. Hence, to obtain the amount of greenhouse emissions CO_{2e} (mass of carbon dioxide equivalent), the mass of greenhouse gas **k** (CO_2, CH_4, N_2O) is multiplied by its corresponding $\mathbf{gwp_k}$ factor (1, 21, 310, respectively), Guinée et al (2002). The emissions of SF_6 and CFCs are negligible in fossil fuel combustion and during electricity life cycle (Dones et al, 2004), thus they were not considered in the present work. Therefore, the life cycle greenhouse emissions for the entire network are calculated as follows:

$$\mathbf{F_{GHG}^{LC}} = \sum_{q \in F} \sum_{f} \sum_{l} \mathbf{G(q,f)} \times \mathbf{E_{GHG}^{l}(q,f)} + \sum_{q \in NF} \sum_{l} \mathbf{G(q)} \times \mathbf{E_{GHG}^{l}(q)} \qquad \mathbf{l = 1,...,l_q} \qquad (12)$$

$$\mathbf{E_{GHG}^{l}(q,f)} = \sum_{k} \mathbf{E_{k}^{l}(q,f)} \times \mathbf{gwp_k} \qquad (13)$$

$$\mathbf{E_{GHG}^{l}(q)} = \sum_{k} \mathbf{E_{k}^{l}(q)} \times \mathbf{gwp_k} \qquad (14)$$

Where $\mathbf{E_k^l}$ is the emission factor of greenhouse gas **k** in the life cycle stage **l** in *ton of CO_{2e}/Gwh*. The subscript **GHG** indicates the summation over the three greenhouse gases considered in each life cycle stage **l**. The total number of life cycle stages, $\mathbf{l_q}$ considered are:

  i. *Thermoelectric power generation*: exploration, extraction, refining, transport and generation step for four different fuels: natural gas, fuel oil, gas oil and coal.
 ii. *Hydroelectric power generation*: material transport in construction and submerged biomass decay in the operation.
iii. *Nuclear power generation*: exploration, extraction, refining, fuel assembly and transport of uranium, waste treatment and disposal of spent fuel, transport in construction phase of the nuclear plant.

A detailed analysis of each life cycle stage considered as well as the literature sources was presented in Eliceche and Martinez (2007).

3. Optimization Problem Formulation

The objective is to determine the optimal configuration and load distribution for all power plants, to provide the electricity demand to the grid minimizing life cycle green house emissions and operating cost simultaneously. A combined objective function includes the operating cost and the benefit for reducing green house emissions. The greenhouse gases international market price $\mathbf{pr_{GHG}}$ (*US\$/ton of CO_{2e}*) was considered to translate an environmental burden into an economic value.

$$\mathbf{EnvEco(x)} = \mathbf{C_G(x,y)} - \left[\mathbf{F_{GHG}^{LC}(x,y)} \Big|_{in} - \mathbf{F_{GHG}^{LC}(x,y)} \Big|_{o} \right] \times \mathbf{pr_{GHG}} \qquad (15)$$

The benefit (term in brackets) is proportional to the reduction in green house emissions from the initial point (sub index **in**) and the optimal solution (sub index **o**). The constant term corresponding to the initial point life cycle greenhouse emissions is removed from Eq. 15 leading to the objective function named *"$\mathbf{EnvEco}$"* in the following Mixed Integer Linear optimization problem P1:

$$\min_{\mathbf{x},\mathbf{y}} \quad \mathbf{EnvEco(x,y)} = \mathbf{C_G(x,y)} + \mathbf{F_{GHG}^{LC}(x,y)} \times \mathbf{pr_{GHG}}$$

$$\mathbf{s.t.}$$
$$\mathbf{A.x} = 0$$
$$\mathbf{B.x} + \mathbf{C.y} \le 0 \qquad\qquad\qquad P1$$
$$\mathbf{x} \in \mathbf{R^n}$$
$$\mathbf{y} \in \{0,1\}$$

Where **x** and **y** are the continuous and discrete optimization variables respectively; equality constraints represent the generation model given in Equations 1 and 2 and the emission model given in Eq. 12 with life cycle emission factors given in Eq. 13 and Eq. 14. Inequality linear constraints are included to represent minimum and maximum plant capacity constraints, demand satisfaction constraint and bounds on continuous variables, Equations 3 to 11. The continuous optimization variables of the problem are the availability factors for all the power generation plant. Fuel resource constraints and mandatory fuel usage can be included as constraints to represent temporal resource availability due to natural gas pipeline capacity and national policies regarding for example the use of coal. It is also possible to include an inequality constraint with a minimum required reduction in total greenhouse emissions to comply with Kyoto protocol (IPCC, 2001) targets stated for a given country or industrial branch. This constraint was not included, because Argentina has no greenhouse emissions reduction targets up to now.

4. Numerical Results

The configuration and loads of the power plants of the Argentinean interconnected electricity system were selected optimally minimizing operating cost and green house emissions as formulated in problem P1. There are 51 thermoelectric power plants consuming 4 alternative different fossil fuels, 41 hydroelectric power stations and 2 nuclear power plants. The 51 thermoelectric power plants have 170 thermoelectric machines including gas and steam turbines and combined cycle units. Data of system operation and installed capacity, availability factors bounds, power plant efficiencies and generation costs were taken from the electricity national company, CAMMESA (2004). The electricity demand for year 2004 was formulated as an inequality constraint, being active at the solution point. The initial point of the optimization problem corresponds to the operating conditions of year 2004. The optimization problem P1 is a mixed integer linear programming problem (MILP) solved in GAMS (Brooke, 2003), using CPLEX as a MILP solver. The main results comparing the operation of the electricity network in year 2004 (the initial point) and the optimal operating conditions (solution point minimizing EnvEco) are shown in Table 1. An international price of U\$S 20/ton CO_{2e} was considered (Point Carbon, 2004).

Significant reductions of 48 % and 44 % in green house emissions and operating cost are reported by properly selecting the configuration and loads of the electricity generator plants. More constraints reinforcing national regulations related to the use of certain fuels like coal in order to keep employment in mining areas, might reduce the improvement reported in this case study.

Thermoelectric generation is reduced, while hydroelectric and nuclear generation is increased. Thermoelectric units that were kept switch on, were those more efficient and less pollutant as it is the case of combined cycle thermoelectric units burning natural

gas. The shut down thermoelectric units were burning fuel oil, gas oil and coal, which are the most expensive and pollutant fossil fuels. Hydroelectric, nuclear and combined cycle burning natural gas are also the cheapest options.

The methodology presented can be extended to different regional applications and time periods.

Table 1. Improvements achieved selecting the power plants in operation.

		Year 2004	Min EnvEco	% Reduction
GHG Emissions	CO_{2e} 10^3 ton	42.00	21.75	48.21
Operating Cost	10^3 US$	2407.67	1353.62	43.77
EnvEco	10^3 US$	3247.69	1788.71	44.92
Thermoelectric	Gwh / plants	42766.00 / 170	38871.16 / 23	9.11
Hydroelectric	Gwh / plants	31827.00 / 41	35019.58 / 40	-10.03
Nuclear	Gwh / plants	7312.90 / 2	8015.16 / 2	- 9.60
Total Electricity Generation Gwh		81905.90	81905.90	0.00

5. Conclusions

The methodology developed for the selection of the operating conditions of the Argentinian electricity network leads to significant reductions of more than 40 % in operating cost and green house emissions simultaneously. Hydroelectric, nuclear and the most efficient and less pollutant thermoelectric units, as the combined cycle burning natural gas, are in operation while the less efficient thermoelectric power plants burning coal, fuel oil and gas oil are shut down. Hydroelectric, nuclear and combined cycle burning natural gas plants are also the cheapest options. Thus, a useful tool to support a decision-making process in a key energy sector has been presented.

References

A. Brooke, D. Kendrick, A. Meeraus, R. Raman (eds.), 2003, GAMS, A user guide, GAMS Development Corporation, Washington DC.

A. Eliceche and P. Martinez, 2007, Minimization of life cycle CO2 emissions in the operation of a steam and power plant, ESCAPE17, V. Plesu and P.S. Agachi (Editors).

CAMMESA, 2004, Compañía Administradora del Mercado Eléctrico Mayorista Sociedad Anonima, Informe Anual 2004. Buenos Aires.

D. Weisser, 2007, A guide to life-cycle greenhouse gas (GHG) emissions from electric supply technologies, Energy, 32, 1553-1559.

H. Hashim, P. Douglas, A. Elkamel and E. Croiset, 2005, Optimization Model for Energy planing with CO2 Emission Consideration, Ind. Eng. Chem. Res., 44, 879-890.

IPCC. Climate Change 2001: the scientific basis. Contribution of working group I to the third assessment report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University. www.ipcc.ch

J. Guinée, R. Heijungs, G. Huppes , R. Kleijn, A. Koning, L. van Oers, A. Sleeswijk, S. Suh, H. Udo de Haes (eds.), 2002, Handbook on Life Cycle Assessment. Operational Guide to the ISO Standards. Kluwer Academic Publishers, Dordrecht.

Point Carbon, 2004. Carbon Market Consulting Services. www.pointcarbon.com

R. Dones, T. Heck and S. Hirschberg, 2004, Greenhouse gas emissions from energy systems, comparison and overview, Encyclopedia Energy, 3, 77-95.

A Systematic Framework to Calculate Economic Value and Environmental Impact of Biorefining Technology

Norman Sammons Jr.[a], Wei Yuan[a], Susilpa Bommareddy[a], Mario R. Eden [a],

Burak Aksoy[b], Harry Cullinan[b]

[a]*Auburn University, Department of Chemical Engineering, 222 Ross Hall, Auburn University, AL 36849-5127, USA, edenmar@auburn.edu*
[b]*Auburn University, Alabama Center for Paper and Bioresource Engineering, 242 Ross Hall, Auburn University, AL 36849-5127, USA, culliht@auburn.edu*

Abstract

The integrated biorefinery has the potential to provide a strong, self-dependent, and sustainable alternative to the use of fossil fuels for the production of chemicals and energy, but difficulties arise in measuring the potential economic benefit and environmental impact of current and emerging biorefinery technology. A plethora of products and production pathways are possible in this growing field of biorefining, and the production path with maximum value and minimum environmental impact cannot be determined by heuristics alone. A framework is needed to determine the most optimal route based on measures of economic and environmental performance. Gross profit is used as a short-term economic metric requiring few details, while net present value is used as a long-term metric that requires information about rate of return, regulatory and market possibilities, and hedging options. Environmental impact is evaluated using the US-EPA WAR algorithm developed by Young and Cabezas (1999). The best candidates in economic and environmental performance are then subject to process integration techniques in order to maximize resource usage, and these integrated biorefineries are once again analyzed for optimal performance.

Keywords: Environmental impact, Economic analysis, Biorefinery, Product allocation, Optimization

1. Introduction

The integrated biorefinery provides a unique opportunity for reinvigorating an entire manufacturing sector by utilizing a renewable resource to create valuable product streams in the forms of chemicals, fuels, pharmaceuticals, and energy (Bridgwater, 2003). Through the optimal use of renewable feedstocks such as forest residues, forest byproducts, and crop residues, long-term economic and environmental sustainability can be achieved. The vast amount of possibilities in feedstocks, chemical processes, and final product routes results in a need for a process systems engineering (PSE) approach to ensure maximum economic and societal benefit through minimizing the usage of raw material and energy resources as well as the cost involved in supply chain operations intrinsic to biorefining. A unique partnership has been established consisting of researchers, government entities, equipment vendors and industry stakeholders to maximize the applicability of such systematic methods and to integrate experimental

and modeling work. This partnership has played an integral role in the procurement of the wide range of information necessary such as data needed for process models, capacity constraints, financial data, and optimization techniques. The overall goal of this work is to develop a system that will enable decision makers to evaluate novel production pathways in biorefining in order to optimize profitability while determining and minimizing environmental impact.

2. Development, Integration, and Evaluation of Process Designs

In biorefining, the large number of possible process configurations and products results in a highly complex problem that cannot be solved using simple heuristics or rules of thumb. Thus, it is necessary to develop a framework which incorporates profitability measures as well as environmental metrics and other techno-economic metrics. This framework should enable policy and business decision makers to answer a number of important questions like:

- For a given set of product prices, what should the process configuration be, i.e. what products should be produced in what amounts?
- For a given product portfolio, how can process integration methods be utilized to optimize the production routes leading to the lowest environmental impact?
- What are the discrete product prices that result in switching between different production schemes, i.e. what market developments or legislative strategies are required to make a certain product or pathway attractive?
- What are the ramifications of changes in supply chain conditions on the optimal process configuration?

The introduction of process system engineering (PSE) methods into biorefining research provides a systematic framework capable of seamlessly interfacing results generated in both experimental work and simulation studies. Such a framework is imperative when attempting to combine the vast array of knowledge and information from many research areas and disciplines. Figure 1 illustrates the data generation and flow necessary for this framework. While the overall strategy of the framework will be reviewed here, complete details of the strategy are presented in earlier work (Sammons et al., 2007).
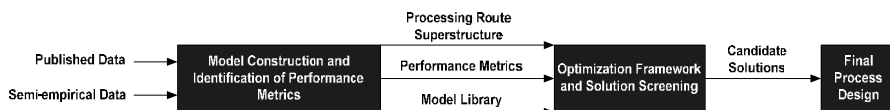
Figure 1 – Overall data generation and flow necessary for product allocation framework.

An initial superstructure lists feasible technologies for a given feedstock, and basic simulation models are then constructed for those corresponding processes based on published and semi-empirical data. If a given model is solvent-based, computer-aided molecular design (CAMD) and property clustering techniques are employed to identify alternative solvents that minimize safety and environmental concerns (Eden et al., 2004; Harper and Gani, 2000). Process integration techniques are then used to optimize the simulation models. This is an integral step in the model development as it ensures optimal utilization of biomass and energy resources, which will inevitably result in reduced environmental impact. The optimized models are used to generate data for the economic as well as environmental performance metrics. The estimation of

environmental performance is achieved through the use of the US-EPA Waste Reduction (WAR) algorithm (Young and Cabezas, 1999). The end result of this strategic portion is a superstructure of biorefining processing routes, a library of simulation models for these routes, and a database of corresponding economic and environmental performance metrics.

As stated in previous work, this superstructure and database is then exported into a mathematical optimization program whose objective is to identify candidate solutions based on maximum profitability (Sammons et al., 2007). The environmental impact of these candidate solutions is also measured for ranking purposes. If a candidate solution satisfies both economic objectives and environmental constraints, then the optimal production pathways have been identified. However, if none of the candidate solutions satisfy the environmental impact requirements, then the framework is modified by relaxing economic performance constraints until one or more solutions will satisfy both criteria. In doing this, economic and environmental performance is decoupled in order to avoid the pitfall of identifying the zero production facility that will identify the minimum environmental impact (Sammons et al., 2007). It should be noted, that only the economic and environmental performance metrics are incorporated in the solution framework described below, thus decoupling the complex models from the decision-making process. This approach allows for updating the model parameters as new data and/or technology becomes available without having to change the selection methodology, thus making it robust and flexible.

3. Generalized Biorefinery Model and Economic Metric Optimization

Figure 2 illustrates the generalized biorefinery model that has been used to develop the structure of the optimization framework. The model structure was formulated to include a variety of basic complexities encountered in the decision making process (Sammons et al., 2007).The objective function maximizing the overall gross profit of the biorefinery over a given time period *t* is illustrated in Equation 1, but this may be replaced with a more robust objective function maximizing net present value as shown in Equation 2.
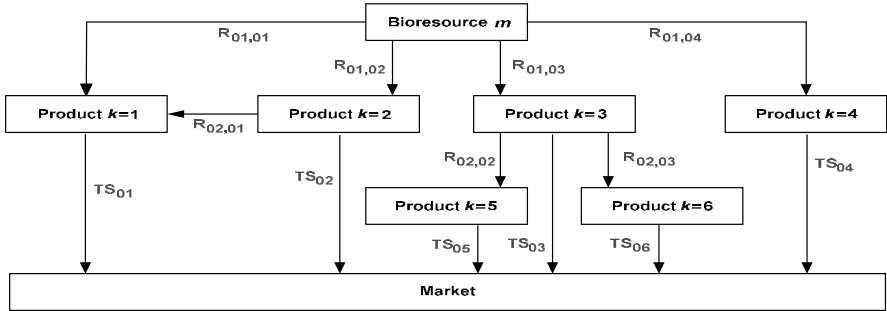
Figure 2 – Generalized model to illustrate possibilities in decision-making tree.

$$GP_t = \sum_m \left(\sum_k TS_{mkt} C_{kt}^s - \sum_i \sum_j R_{mijt} C_{mijt}^P - C_{mt}^{BM} \sum_j R_{m1jt} \right) \quad (1)$$

$$\max NPV = \sum_t \frac{GP_t(1-Tax_t) + Dep_t Tax_t - Hedge_t + Gov_t}{(1+R)^t} \quad (2)$$

Using this nomenclature, the first set of terms in Eq. (1) represents the sales revenue from the products made from each bioresource m. TS_{mkt} is a variable that denotes the production rate of product k from bioresource m that is sold to the market during time period t. C_{kt}^s is the sales price of product k, which is a scalar and is determined through a market survey of published prices and vendor quotes. The second set of terms represents the total processing cost incurred by the pathways pursued in production. R_{mijt} is a variable that represents the processing rate of route ij while C_{mijt}^P is a scalar that represents the processing cost of that route and is determined through simulation models and process economics. The third set of terms represents the total cost of the biomass resource m, and this is broken down into the scalar purchase price of bioresource m in C_{mt}^{BM} and the combined rate of biomass processed by the plant in R_{m1jt}. Although both TS_{mkt} and R_{mijt} are variables in the optimization program, they are not independent since the variables are related via mass balance constraints around the product points and process conversion factors.

This gross profit term may be maximized by itself or incorporated into a net present value objective function that takes many more issues into account. Tax_t represents the marginal tax rate, $Hedge_t$ represents the expenses associated with hedging against catastrophic market actions, Gov_t is the net benefit realized through government incentives or penalties, and R is the expected rate of return, or cost of capital involved with the time value of money. The gross profit version of the model may be utilized for short-term decision making and scheduling with existing equipment and supply chain arrangements, while the net present value version is better suited for long-term planning that will involve extensive capital expenditures, anticipated government actions, and supply chain considerations.

4. Results and Case Studies

Without including any constraints on capacity of the process points, the solution is a single-product configuration in which all available biomass is converted into the most profitable product, or the product with the greatest contribution margin. However, if constraints are imposed on the most profitable route, the framework identifies the additional products and processing routes with the next highest contribution margins in order to maximize profitability, thus leading to a polygeneration facility (Sahinidis et al., 1989). In order to effectively address the strategic planning objectives of business decision makers, it is necessary to incorporate the total capital investment as a constraint in the formulation. Inclusion of capital cost constraints is crucial for practical application of the results, i.e. enabling consistent evaluation of the potential benefits to be obtained for a given maximum investment.

A case study was performed on a potential grassroots biorefinery involving the conversion of chicken litter to syngas, hydrogen, and electricity. Conversion rates were obtained from experimental work performed by the university and affiliated agencies as well as simulations constructed in ASPEN. Figure 3a shows the possible pathways for production and sale of these products on the common market. The execution of the framework verified the results obtained from manual calculation; producing syngas from chicken litter and selling it on the market via pipeline would maximize profit due to the high costs intrinsic with converting the syngas to hydrogen or electricity. Figure 3b highlights the active pathway chosen by the optimization program.
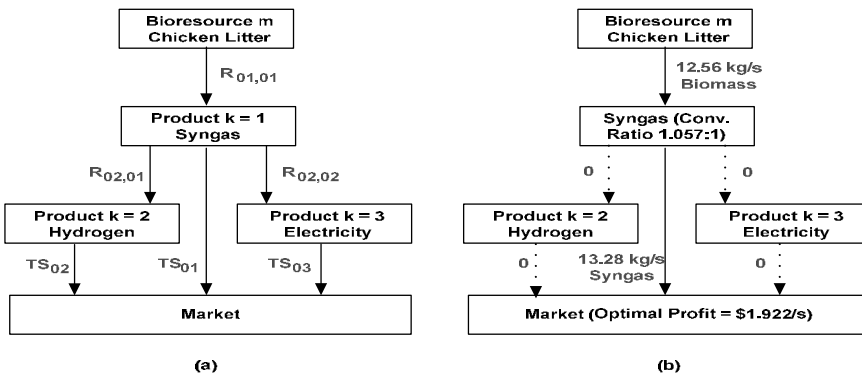
Figure 3 – (a) Unsolved decision making tree. (b) Solved decision making tree with flowrate values and objective function value.

A second case study was performed on the potential of gasification of black liquor byproduct from pulp and paper mills and its subsequent conversion into electricity, steam, and liquid fuels. These processes were compared with the use of a replacement state-of-the-art Tomlinson boiler, which represents the status quo of burning the organic components of black liquor to generate steam in a relatively inefficient manner. Data from an extensive study performed by Larson et al. (2006) was used to construct simulation models that provided the data necessary for the framework to be utilized. Figure 4 illustrates the unsolved decision making tree for the black liquor case study, which shows the process models simulated and observed in the Larson study. From execution of the optimization code, the framework suggested that all biomass should be gasified and processed through the FTc pathway, which represents one of the ways in which Fischer-Tropsch liquid fuels may be produced through the gasification of black liquor along with gasifying supplemental biomass. This result, shown in Figure 5, is qualitatively confirmed by the observation made by Larson et al. (2006) that this process is the most profitable since it has the highest net present value over time.
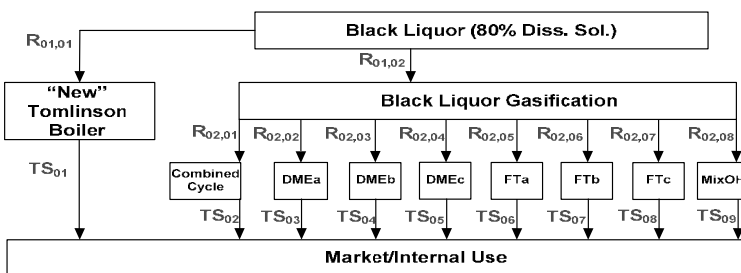
Figure 4 – Unsolved decision making tree for black liquor gasification study based on data from Larson et al (2006).

5. Conclusions and Future Work

A general systematic framework for optimizing product portfolio and process configuration in integrated biorefineries has been presented. Decoupling the process models from the decision-making framework reduces problem complexity and increases robustness. The next phase of this work involves development of additional process

models for the generation of performance metrics, specifically information on conversion, yield, and production cost for economic metrics.
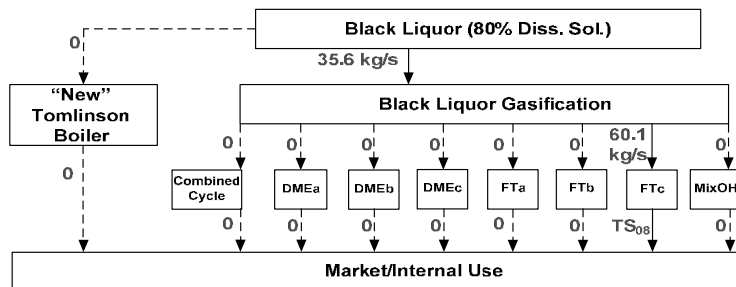
Figure 5 – Solved decision making tree for black liquor gasification study. TS_{08} contains multiple products of FT fuels, electricity, and process steam.

The EPA WAR algorithm will be incorporated into the mathematical optimization software, and data from the additional process models will be used to generate numerous time-based and mass-based measures of environmental impact. From there, process integration will be utilized to optimize these process models by reducing energy usage, material consumption, and waste streams. The framework will also become a stronger financial tool through the incorporation of various economic ideas and analyses. The further development of qualitative predictive models for capital investment and inclusion of capital amortization into the objective function will also increase the strength of the framework. Incorporation of options theory into the framework will allow management to develop financial strategies in response to events in the market or legislative environment. Optimization under uncertainty and superstructure generation techniques will be studied to increase the robustness of the framework (Banerjee and Ierapetritou, 2003; Chakraborty and Linninger, 2003).

References

A.V. Bridgwater, (2003). Renewable fuels and chemicals by thermal processing of biomass. Chemical Engineering Journal, 91, 87-102.

M. R. Eden, S. B. Jørgensen, R. Gani and M. M. El-Halwagi, (2004). A novel framework for simultaneous separation process and product design. Chemical Engineering and Processing, 43, 595-608.

P. M. Harper and R. Gani, (2000). A multi-step and multi-level approach for computer aided molecular design. Computers and Chemical Engineering, 24, 677-683.

D. M. Young and H. Cabezas, (1999). Designing sustainable processes with simulation: the waste reduction (WAR) algorithm. Computers and Chemical Engineering, 23, 1477-1491.

N. E. Sammons, W. Yuan, M. R. Eden, H. T. Cullinan, B. Aksoy, (2007). A flexible framework for optimal biorefinery product allocation. Journal of Environmental Progress, 26(4), 349-354.

N. V. Sahinidis, I. E. Grossmann, R. E. Fornari and M. Chathrathi, (1989). Optimization model for long range planning in the chemical industry. Computers and Chemical Engineering, 13, 1049-1063.

E. D. Larson, S. Consonni, R. E. Katofsky, K. Iisa, and W. F. Frederick, (2006). A Cost-Benefit Assessment of Gasification-Based Biorefining in the Kraft Pulp and Paper Industry, Volume 1, Princeton University.

I. Banerjee and M. G. Ierapetritou, (2003). Parametric process synthesis for general nonlinear models. Computers and Chemical Engineering, 27, 1499-1512.

A. Chakraborty and A.A. Linninger, (2003). Plant-wide waste management 2: Decision making under uncertainty. Industrial and Engineering Chemical Research, 42, 357-369.

Development of a Computer Support System for the Management of Regulatory Compliance of Pharmaceutical Processes

M. Berkan Sesen[a], Pradeep Suresh[b], René Bañares-Alcántara[a],*, Venkat Venkatasubramanian[b]

[a]*Department of Engineering Science. University of Oxford, Oxford OX1 3JP, UK*
[b] *School of Chemical Engineering. Purdue University, W. Lafayette, IN 47907, USA*

Abstract

The pharmaceutical sector is one of the most tightly regulated industries today and is constantly being challenged to meet rising standards of quality. However, it still uses paper documents, spreadsheets and conventional databases for the storage and manipulation of the manufacturing and regulatory process knowledge. Furthermore, in the current industrial approach, the interpretation of the regulations (which are written at a very abstract level) into operating procedures is done manually and as an afterthought to pharmaceutical process development. This approach is error-prone, time consuming and very effort intensive as it does not take advantage of recent advances in the field of knowledge management.

We have been working in the development of a computer-based support system to assist in the identification of regulatory compliance of a drug manufacturing process. OntoReg, the current prototype, encapsulates pharmaceutical process and regulation knowledge in two complementing representations:

1. OWL ontologies (a knowledge representation consisting of taxonomies of concepts and logical axioms allowing to structure those concepts and to detect inconsistencies in the resulting structure), and
2. SWRL rules (which act as constraints able to enforce values inside the concepts or create relations between them).

These two components, ontologies and rules, are integrated through a Java user interface which is able to identify when a pharmaceutical process does not comply with a regulation and to suggest remedial action.

The resulting OWL ontology is structured in terms of three types of concepts belonging to a Regulatory, a Process or an Abstract domain (such as Time and Parameter). OntoReg has been tested with a case study for the aspirin production process in the context of its compliance with some equipment maintenance and cleaning regulations taken from the European Union Guidance on Good Manufacturing Practice.

Our approach has the potential to substantially decrease pharmaceutical process validation time (up to several weeks) and effort and thus reduce development costs and commercialization prices (a well known estimate is $1 million per day profit for a blockbuster drug). Furthermore, this novel approach has the potential to be extended to other regulatory applications in the future, e.g. environmental compliance.

Keywords: Decision support systems, Pharmaceutical processes; Regulatory Compliance; Ontologies.

1. Introduction

During the preparation for pharmaceutical process validation, data must be systematically gathered, analyzed and presented to give reasonable documented evidence that a pharmaceutical process, when operating within specified parameters, will "consistently" produce a product that meets its pre-specified quality attributes. Closely related to both are Compliance Management Systems (CMSs), i.e. systems that keep track of the tasks needed to comply with the validation requirements. Conventionally, compliance management is a manual and highly labor-intensive task that creates additional overheads to many businesses [1]. Pharmaceutical manufacture, in particular, is one of the most tightly regulated industries today and it is constantly being challenged to meet rising standards of quality and to comply with rigorous regulatory requirements. Ensuring regulatory compliance and managing a myriad of validation documents constitutes a major engineering challenge because manual handling of voluminous and heterogeneous data is not only time-consuming but also very expensive and error-prone.

 In the context of the pharmaceutical industry, while front end applications have been designed to automate process controls for drug manufacture, back end data collection and storage applications have significantly lagged behind, still relying on manual and paper based archiving methods. Motivated by this need, various compliance management systems have been offered in the past. In 2000, when electronic record keeping was still new to the Food and Drug Administration, Conley [2] looked at automating regulatory compliance systems, focusing mainly on the NuGenesis Scientific Data Management System (SDMS). In 2006, Boland et al. [3] compared the key features of the main commercial CMSs within the context of environmental management. Yip et al. [1] used semantic modeling to develop a compliance audit mechanism, namely XISSF, which enforces information security standards and regulations. These important initial attempts, while recognizing the need for an informatics framework to automate CMSs, are limited in their conceptualization of the raw regulatory data, parsing of regulations into tasks and in their application to the pharmaceutical domain. In terms of compliance systems, establishing the correlations and relationships among raw data in order to acquire useful business information is vital. However, a systematic way to convert the raw/implicit regulatory data to human-interpretable and machine-readable explicit information is still lacking.

The rest of the paper is organized as follows: Section 2 presents the methodology followed to develop the OntoReg platform, its architecture and structure. Results are presented in Section 3, and Section 4 concludes with the contributions and future work.

2. Methodology

Motivated by the lack of an existing tool, this work aims to prove that an ontological knowledge-base infrastructure can be applied to the pharmaceutical process regulation domain in order to facilitate the management of the validation requirements. In a nutshell, ontologies capture and describe the semantics of data sources and make their contents explicit and platform-independent. They are not only useful for unifying database, data-warehouse, and knowledge-base vocabularies but also for maintaining consistency through inferential reasoning tools. Compared to a database schema (which targets physical data independence) and to an XML schema (which targets document structure), an ontology targets shared and explicit semantics of information [4]. In our application, this explicit semantic knowledge is embedded on a domain specific ontology, named OntoReg, and processed by the combination of an ontological reasoner

(Pellet 1.5.2) and a rule engine (Jess 7.1) through the use of a Java integrated development environment (IDE). The ontology aims to capture pharmaceutical regulatory and process domain knowledge in a generic way and to provide a common representation of the domain, which may be reused, shared, and standardized across applications and groups [5]. The contribution of this work is the use of this domain specific ontology, i.e. OntoReg, to form an integrative platform, which amalgamates rule enforcement and knowledge inference functionalities through a Java application.

OntoReg was initially developed on paper and then formalized in the OWL language with Protégé–OWL graphical editor. The key objective behind its design was to formalize the tacit human concepts in a way that would adequately represent physical (which is subdivided into regulatory and process domains) and abstract concepts and allow explicitly formal reasoning, which is both understandable by humans and processable by software, see Figure 1.
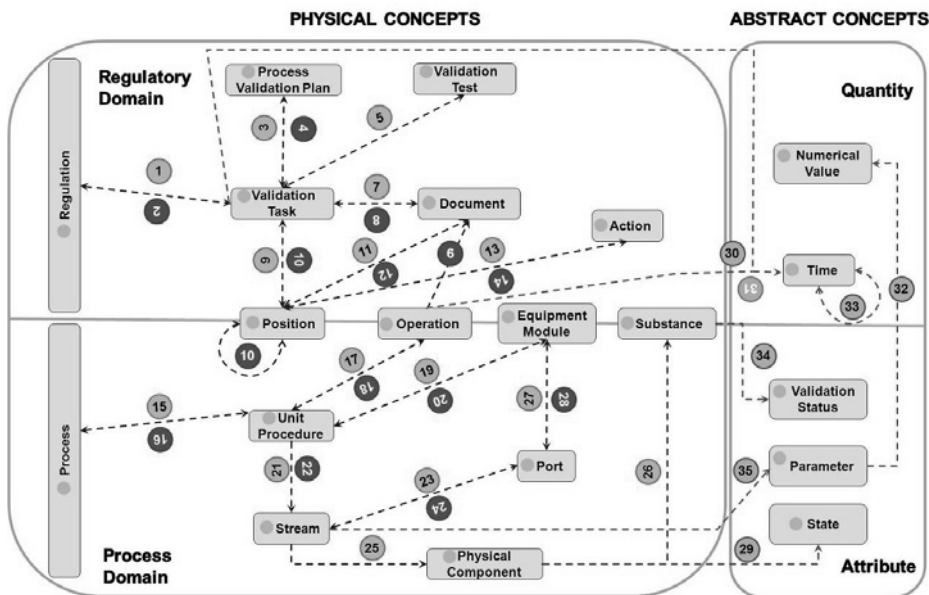
Figure 1. The main classes and interrelations (T-Box) in OntoReg.

A few process engineering ontologies, such as OntoCAPE [6] and POPE [7], were studied to capture relevant concepts that would help in the development of the Process Domain of OntoReg. Additionally one upper class ontology, namely SUMO [8], was examined but eventually deemed to be incompatible with description logics (DL) reasoning that is inherent to the OWL-DL language.

Figure 2 depicts how the ontological knowledge base OntoReg, lying at the heart of the whole infrastructure, is utilized within the application. The application is organized in three layers. The Knowledge Layer, where the regulatory and process concepts are encoded into the ontology in the form of classes, axioms and rules (constituting the Terminological Box or T-Box). The Inference Layer, where the rule engine is used to interpret and enforce the rules imposed on the ontological knowledge base. In addition, a Description Logics reasoner is also present in this layer and is used to classify the taxonomy and compute inferred individual types that are asserted by axioms. The third layer is the Java Integration Layer, which integrates the first two layers along with

presenting them through a front end application for user interaction. The Java engine and GUI (Graphical User Interface) of the OntoReg platform are designed to aid the user in the process of resolving logical inconsistencies and furnishing critical missing data in the knowledge base. All these tools interact with each other to provide an integrated environment that allows deduction of new facts while constraining existing ones in order to ensure regulation compliance of a pharmaceutical process.
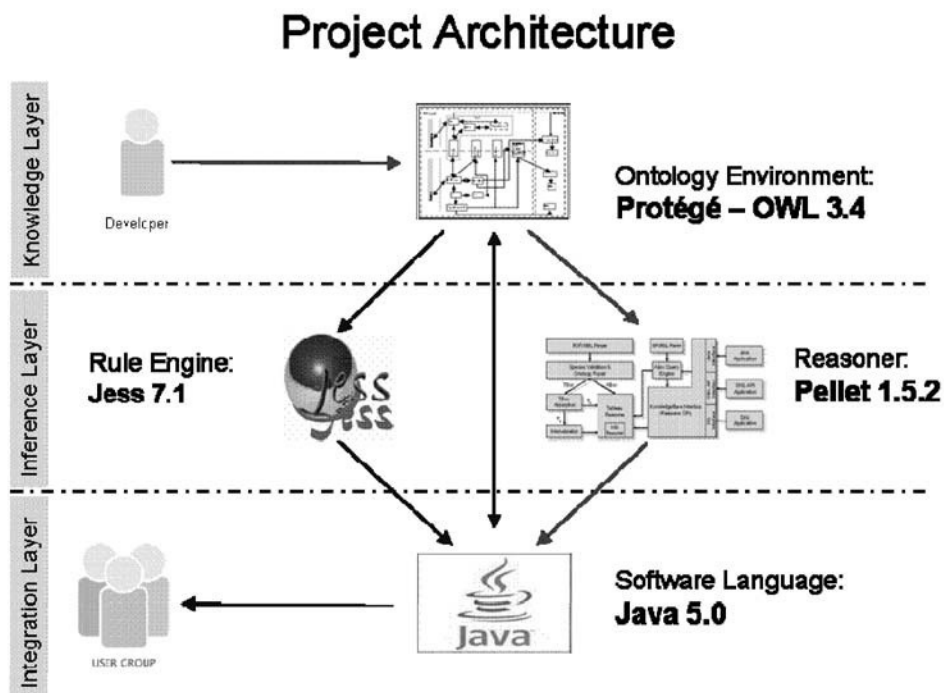
Figure 2. Architecture of the OntoReg platform.

3. Results

The structure of the regulatory domain in OntoReg has been derived after carefully going through the European Union Drug Regulatory Network Legislation (Eudralex) Guide 2007 [9] and Haider's Pharmaceutical Validation Master Plan [10]. After the conceptualization of the ontology was complete, a few small portions of the Eudralex guidelines were selected (varying from Cleaning Validation to Process Validation and Substance Qualification) for the creation of these case studies that demonstrate the capabilities of the current platform in the enforcement of the given regulations on an Aspirin production process [11] introduced to the knowledge base. Discussed below is one simple case study that shows the use of the OntoReg platform and the value added towards automated regulatory compliance. This particular example shows how OWL axioms can be used to enforce some basic requirements such as the need that a process material is to be supplied by a vendor who is approved by the organization (the pharmaceutical company). The axioms involved in enforcing this requirement are

1 `StartingMaterial isSuppliedBy` <u>`some`</u> `(PharmaSupplier` <u>`and`</u>
`(hasSupplierAssessment has Approved))`

2 `StartingMaterial isSuppliedBy` <u>`exactly`</u> `1`

3 `PharmaSupplierAsssess isResponsibilityOf` <u>`some`</u> `(PurchasingOfficer` `and` `(performs` <u>`only`</u> `Assessing))`
4 `PharmaSupplierAsssess hasReport` <u>`some`</u> `SupplierAssessmentReport`
5 `PharmaSupplierAsssess hasReport` <u>`exactly`</u> `1`
6 `ApprovedSupplier hasSupplierAssessment` <u>`has`</u> `Approved`
7 `DisapprovedSupplier hasSupplierAssessment` <u>`has`</u> `Disapproved`
8 `ApprovedSupplier isPatientOf` <u>`some`</u> `PharmaSupplierAssessTask`

Now, to start with, assume that salicylic acid (which is used in the Aspirin production process) has no associated vendor or supplier. This would lead to the warning shown below indicating that StartingMaterial *Salicylic_Acid* does not have an "isSuppliedBy" relation with any Suppliers (violates Axiom2), see Figure 3.

Figure 3. The Missing Data warning: a supplier has to be assigned for Salicylic_Acid.

In the next step the user sees the cardinality violation and chooses *CBAChemicals* as the supplier of *Salicylic_Acid*. Assume that *CBAChemicals* is an individual of the DisapprovedSupplier class (violates Axiom1). After selecting this supplier, and on running a consistency check, the reasoner displays an error message stating that *CBAChemicals* can not be the supplier of a StartingMaterial (not shown).

The user realizes his/her mistake and finds that the supplier of *Salicylic_Acid* was indeed *ABCChemicals*, an ApprovedSupplier, and not *CBAChemicals*; which fixes this validation problem.

Further case studies, which are more lengthy and complicated and that make use of SWRL rules, can be found at [12] and show the real power of an automated system and the amount of regulatory burden it can circumvent. It should be noted though that these case studies were chosen to be merely proofs of concept studies to showcase the capabilities of a semantic approach towards an automated process validation framework

4. Conclusions and Future Work

The majority of the pharmaceutical industry still uses paper documents, spreadsheets and conventional databases for process and regulatory knowledge storage and validation. This approach is error prone and does not take advantage of the advances in the field of information technology and knowledge management. This project proposes an ontological knowledge base, OntoReg, which encapsulates the process and regulation knowledge in OWL ontologies, and uses axioms and rules as means to enforce

pharmaceutical regulatory guidelines on the drug manufacturing process. This approach treats regulatory compliance as a parallel activity to the development of drug manufacturing process rather than as an afterthought, ensuring rapid elimination of gross oversights and inconsistencies that could prove costly once the drug making process is under review from the regulators. By partly automating the match between regulatory guidelines and a pharmaceutical process, and reporting any inconsistencies, this approach also avoids human errors. Furthermore, the knowledge-based assistance enables a more structured approach to regulatory compliance, and makes assumptions explicit. Valuable know-how, previously residing only in an expert's mind, can be preserved. Overall, a knowledge-based approach may catalyze the whole effort, decreasing compliance time and costs.

Aside from the obvious step of extending the scope of the OntoReg ontology to account for a larger set of principles from the EU Guidance on GMP [9], a number of extensions are currently being implemented in our group:

- Natural language processing to parse regulations and extract knowledge in the form of ontologies and rules.
- Provision of the OntoReg functionality as a set of web services.
- Tailoring OntoReg's functionality in terms of two sets of users: engineers from a pharmaceutical company and validation experts working for a regulatory agency.

Other interesting pieces of future work, such as the checking of the internal consistency of a set of regulations and/or between different sets of regulations (e.g. FDA vs. MHRA) have been pointed out to be important by the pharmaceutical and other regulated industries. In order for OntoReg to evolve, it is crucial for it to be applied to real industrial processes.

References

[1] F. Yip, P. Ray, N. Paramesh, 2006, "Enforcing Business Rules and Information Security Policies through Compliance Audits", in BDIM '06 International Workshop, pp. 81- 90.

[2] J. Conley, 2000, "Automating Regulatory Compliance", Sci. Computing & Instrumentation.

[3] R.F. Boland, 2006, "Reduce Business Risk with a CMS", Chemical Engineering Prog, 102, 10.

[4] V. Venkatasubramanian, C. Zhao, G. Joglekar, A. Jain, L. Hailemariam, P. Suresh, P. Akkisetty, K. Morris, G. V. Reklaitis, 2006, "Ontological informatics infrastructure for pharmaceutical product development and manufacturing", Comps&ChemEng, pp. 1482–1496.

[5] Y. Sure, J. Angele, S. Staab, 2002, "OntoEdit: Guiding Ontology Development by Methodology and Inferencing", in Proceedings of ODBASE.

[6] A. Wiesner, J. Morbach, W. Marquardt, 2007, "An overview on OntoCAPE and its latest applications" AIChE Annual Meeting.

[7] L. Hailemariam, P. Suresh, P. Akkisetty, G. Joglekar, S.H. Hsu, A. Jain, K.R. Morris, G.V. Rekaitis, P. Basu, V. Venkatasubramanian, 2008, "Excipient Interaction Prediction: Application of the Purdue Ontology for Pharmaceutical Engineering (POPE)", in ESCAPE-18 proceedings, Lyon, France.

[8] Suggested Upper Merged Ontology (SUMO), 2008, http://www.ontologyportal.org/.

[9] Medicines and Healthcare products Regulatory Agency (MHRA), 2007, "Rules and Guidance for Pharmaceutical Manufacturers and Distributors".

[10] S. I. Haider, 2002, "Pharmaceutical Master Validation Plan: The Ultimate Guide to FDA, GMP, and GLP Compliance".

[11] E. E.A. Handel-Vega, 2001, "Synthetic Procedure for the Manufacture of Aspirin." vol. US 6278014 B1, U. S. P. Organization.

[12] M.B. Sesen "OntoReg: A Support System for the Management of Regulatory Compliance of Biopharmaceutical Processes", 2008, MSc dissertation, Dept. of Eng. Sci., Univ. of Oxford.

Multi-Objective Game Models for Chemical Industrial Park

Jia Xiao-Ping[a,c], Wang Fang[b], Xiang Shu-Guang[a]

[a]*Center for Chemical Process Engineering, Qingdao University of Science and Technology, Qingdao, 266042, China. E-mail address: cnjiaxp@yahoo.com.cn*
[b]*College of Environmental and Safety Engineering, Qingdao University of Science and Technology,Qingdao, 266042, China.*
[c]*The Guangdong Provincial Laboratory of Green Chemical Technology, School of Chemistry and Chemical Engineering, South China University of Technology, Guangzhou, 510640*

Abstract

Chemical industrial park is a practical form of industrial symbiosis network to emerge to achieve the goal of resource sharing among the participating chemical plants. The aim of this work is to present a systematic framework for industrial symbiosis networks (ISN) evaluation and modeling. A mathematical multi-objective game model is proposed to enhance understanding of ISN issues and motivate improvements of economic, resources and environmental sustainability. Results suggest that the effectiveness of policies designed to improve chemical industrial park management may be enhanced by this proposed framework. It allows the system designers to consider simultaneously the synthesis of ISN according to the existing data to establish synergies between each two participating plants.

Keywords: Sustainability, Game theory, Multi-objective programming, Industrial symbiosis

1. Introduction

Sustainability of industrial systems is becoming increasingly important. To improve economic and environmental performance of industrial system, the members of this system discover the opportunities for internal recycle and external mass or energy exchange and exploring market opportunities for waste [1]. Industrial symbiosis, an important concept and practical application in the field of industrial ecology, is about identifying and using such synergies and linkages between members. It engages traditionally separate industries in a collective approach to competitive advantage involving physical exchange of materials, energy, water, and/or by products [2]. From the theoretical point of view, ISN are also called joint production processes or multi-plant chains (webs). Tools for systematic designing and evaluating such industrial ecology networks are still in their infancy [3]. Therefore, the goal of this work is to present a systematic framework for IS evaluation and modeling.

2. Multi-objective modeling

2.1. Conceptual structure

Each plant considers its resource needs and waste aspects separately in the traditional industrial systems. However, separate plants could be linked through resource sharing

from a practical point of view. ISN provides highly inter-dependent relationship between two plants, exchanging materials and energy in a mutually advantageous manner [4]. Fig.1 shows the conceptual structure of ISN. It should be noted that this structure is addressed by considering a large redundancy of plants and connections. It allows the designers to consider simultaneously the synthesis of ISN according to the existing data collected by the designers seeking to establish synergies with other participating plants.

ISN would provide one or more of the following benefits across inter-plants: 1) reduction in the use of raw materials as resource inputs; 2) reduction in pollution; 3) increased systemic energy efficiency leading to reduced systemic energy use; 4) reduction in the volume of waste products requiring disposal; 5) increase in the amount and types of process outputs that have market value [5].
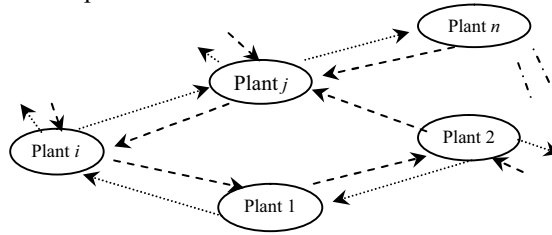
Fig. 1. The conceptual structure of industrial symbiosis network

2.2. Mathematical model

A multi-objective model is proposed mathematically in this subsection. The objectives involve economic, resources and environmental sustainability. Economic objective relates to indicators including investment, net profit, net present value, operation and maintenance cost, and environmental cost. Resource sustainability includes energy, water, and raw materials input, mass of recycled materials, energy and water, and mass of by-products exchanged. Environment objective includes potential environmental impacts associated with releases (including global warming potential, ozone depletion potential, eutrophication potential, photochemical oxidation potential, acidification potential, and human toxicity potential), mass of air and water pollutant released, and mass of solid waste disposed. Each objective is obtained from the combination of indicators using analytic hierarchy process respectively. Thus, the multi-objectives model problem is formulated as shown in the following.

$$\max \quad (f_{\mathrm{ECO}} - f_{\mathrm{RES}} - f_{\mathrm{ENV}})$$

$$s.t. \quad P\,(m_{\mathrm{I}} - m_{\mathrm{O}}) = 0 \qquad\qquad\qquad (1)$$

$$E\,(e_{\mathrm{I}} - e_{\mathrm{O}}) = 0 \qquad\qquad\qquad (2)$$

$$R^k = A^k W^k \qquad\qquad\qquad (3)$$

$$A^k = (a_{ij}^{\ k})_{n \times n} \qquad k, i, j = 1, 2, \ldots., n \qquad (4)$$

where f_{ECO}, f_{RES}, f_{ENV} represent economic, resources, environmental sustainability objectives respectively; P, E represent mass and energy transformation matrices respectively; m, e represent mass and energy matrices of ISN respectively; Subscripts I, O mean inputs and outputs of the members respectively. The system of equations (1~2) represents the mass and energy balances of ISN. In equations (3~4), R means by-products or energy accepted from plant i to plant j. W means by-product or energy

supplied from from plant j to plant i. A means transformation matrix of plant k between feed and by-product or energy discharged.

3. Multi-objective game model

3.1. multi-objective game problem
The members in the ISN are independent because of market economics. The interest and production strategy of an individual member are also independent. Therefore, there are two or more decision makers in this multi-objective situation described in section 2. The strategies of the members are the responses to other members, possibly conflicting each other. It increases the complexity of ISN. Game theory is introduced for solving this problem [6]. Pareto Nash-Equilibrium solution will be obtained. The following subsection will provide a procedure for this multi-objective game problem under discussion.

3.2. game theory with goal programming
A hybrid multi-objective model approach is proposed to solve such problem. It consists of ISN (outer) and separate plant (inner) levels. Game theory is used at the outer level. Game theory with goal programming is used at the inner level. This approach provides an environment between multi-objective problems and decision-makers to incorporate different types of interaction. The stepwise approach is presented (see Fig. 2).

Step 1: Construct ISN problem to find possible alternatives;

Step 2: Construct game theory model G using the following procedure. There are n players Y_i (i=1, 2, …, n). Each plant has its own production strategies S_s (s=1, 2, … , m); the corresponding payoffs are F_i (i=1,2, …, n), for example, F_1 (f_{ECO}, f_{RES}, f_{ENV}) is the payoffs of plant 1.

Step 3: Find the Nash-equilibrium using linear programming.

Step 4: Construct goal programming problem in the inner level such as

$$G(Y', S', F', g, Q_i, w) \tag{5}$$

$$\min \quad Z_i = Q_{il} \sum_{j=1}^{k_i} (w_{ilj}^- d_{ij}^- + w_{ilj}^+ d_{ij}^+), \quad l = 1, 2, \mathrm{L} \ \ L$$

$$F'_{ij}(s_1, s_2, \mathrm{L} \ s_n) + d_{ij}^- + d_{ij}^+ = g_i^j,$$

$$d_{ij}^-, d_{ij}^+, s_n \geq 0$$

$$s_i \in S'_i \quad i = 1, 2, \mathrm{L} \ m$$

Where g, Q and w represent goal constraints, the preference of objective, and goal deviations respectively.

Step 5: Implement goal programming for each objective of each plant and obtain the intersection set C of solution C_i.

Step 6: If C is null set, then minimum Q_i are deleted and return. Else if C exists, then C is determined as Pareto Nash-equilibrium solution.

Step 7: Optimize the objectives of each plant and calculate the objectives of ISN.

Step 8: Decision-making of the production strategies of each member in ISN.
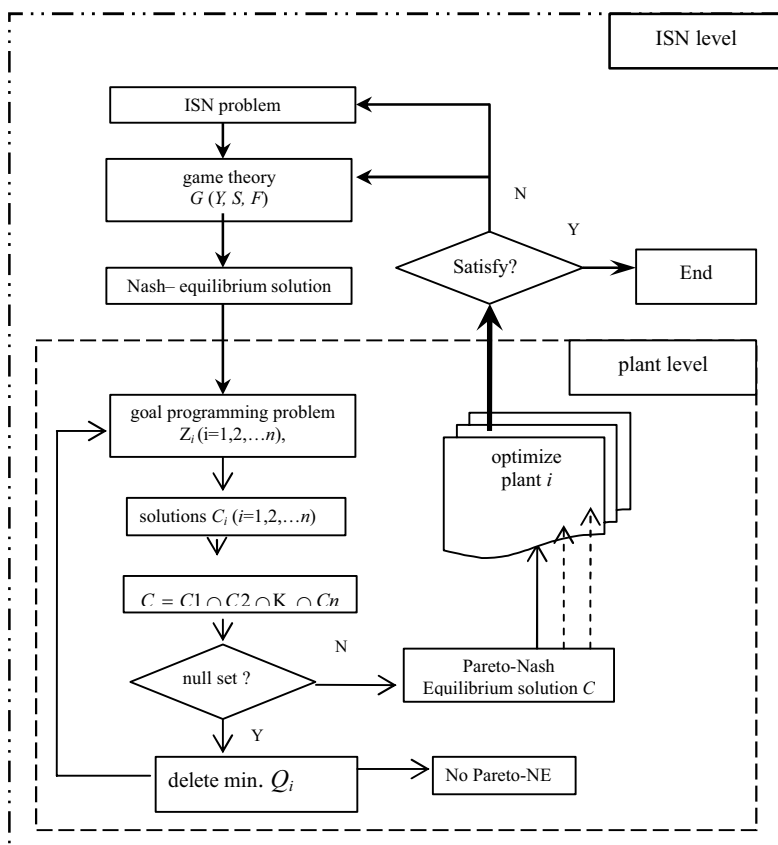
Fig.2. Framework for a multi-objective game problem solving procedure

4. Case Study

4.1. Network Description

This section discusses the symbiotic relationship among four independent companies in a chemical park. Their collaboration represents inter-company exchange of sulphur by-products.

Chemical company A is a complex of basic chemicals and chemical fertilizer. Because of using low level equipments and technologies, large quantities of waste are generated and emitted into air, water and soil. SO_2 emission from boiler flue gas and off-gases of sulfuric acid plant is the most important pollutant. In chemical company C, large quantities of low concentration SO_2 in off-gases are produced from metal sulphide roasters, sulfuric acid plant and lead sintering machines. A large quantity of high concentration sulfur slag is also generated. It has two pyrite-based sulfuric acid plants using double-conversion and double-absorption process. Chemical company D is the off-gas centralized treatment plant.

4.2. Optimal network

Due to high environmental awareness and business opportunities, these four companies will generate new revenues and cost savings for the companies involved and reduced SO_2 pollution to air, water, and land. The new revenues cover SO_2 emission reduction

fee and high value added sulfur products. Based on the materials exchanges among companies, inter-companies sulfur-based flow optimal diagram is shown in Figure 3. Thanks to a limited extent, we present initial results for multi-objective game models only rather than the detailed problem solving procedure in this paper.

4.2.1. PA: Chemical company A

To recover sulphur by-product produced from PC, PA will build a 200,000 tonnes per year sulphur based sulfuric acid plant using double-conversion and double-absorption process to replace the existing pyrite-base one. It will also reduce the off-gas emission, which is sent to off-gas centralized treatment plant. A part of sulfuric acid product is sent to company B to produce sodium hydrosulfite. The SO_2 from the boiler flue gas and off-gases of sulfuric acid plant is recovered, and then used as the raw materials in company B. A part of sulfuric acid is sent to sulfuric acid-based titanium dioxide production process.

4.2.2. PB: Chemical company B

SO_2 recovered from off-gas centralized treatment plant is used as the raw material in PB. SO_2 from the boiler flue gas off-gas in PB is recovered as the raw material. Sulphur from PC is purified and used as the raw material to produce sodium hydrosulfite.

4.2.3. PC: Chemical company C

To fulfill PC's requirements for compliance with sulphur emission standards and profits, low concentration SO_2 is separated into three parts, as shown in Figure 3. In the PC Company itself, Dynawave scrubber gas cleaning technology and Topsoe's wet gas sulfuric acid technology (WSA) are used to recover sulphur and produce commercial-grade concentrated sulfuric acid. The off-gas from WSA plant is sent to a final centralize treatment plant. The second part is sent to PA for producing sulfuric acid. The rest of low concentration SO_2 is sent to PB for producing sodium hydrosulfite. High concentration sulfur slag is used to produce sulphur, which then is sent to PB for producing sodium hydrosulfite.

4.3. Results

The total SO_2 emission reduction is up to 15,000 tonnes per year. The estimated output of industrial symbiosis network in the park is: 0.17 million tonnes of sulfuric acid, 60,000 tons of sodium hydrosulfite, 56,000 tonnes of sulphur.
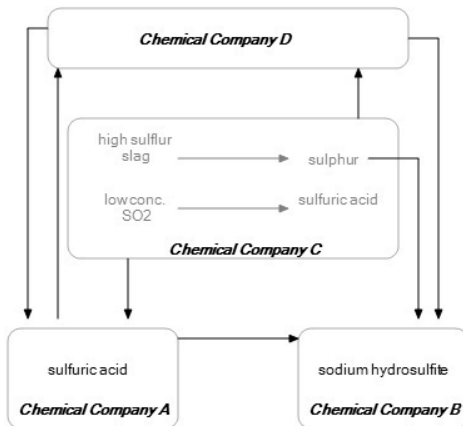
Fig.3 Final network for sulfur-based industrial symbiosis

5. Conclusions

This paper introduced the concept and structure industrial symbiosis network (ISN) to improve the sustainability of industrial systems. The framework for ISN evaluation and modeling is presented. To solve the multi-objective game model problem, a two-level, hybrid multi-objective approach is proposed. The development of ISN offers a potential regional platform to contribute to ensuring efficient utilization resources while minimizing the negative impacts on the environment. To illustrate the concept and potential application, this paper introduced sulfur oriented ISN as the case study. It consists of interlinked production of sulfur related products. These production networks utilize nearly all sulfur by-products to generate new revenues and reduce environmental emissions and disposal costs. It is expected that this approach provides an interactive environment between problems and decision-makers for modeling of industrial systems in the future research.

References

Lou Helen, M. A. Kulkarni, et al., A game theory based approach for emergy analysis of industrial ecosystem under uncertainty, Clean Technologies and Environmental Policy, 6(2004)156

Marian R. Chertow, Industrial symbiosis: Literature and Taxonomy, Annu. Rev. Energy Environ., 25(2000)313

Allen, D. T., R. S. Butner, Industrial Ecology: A Chemical Engineering Challenge, Chem. Eng. Prog., 98(11)(2002) 40

Manahan, Stanley, Industrial Ecology: Environmental Chemistry and Hazardous Waste, Lewis Publishers, Washington D.C., 1999

Gertler,N., Industrial Ecosystems: Developing sustainable industrial structures. Master thesis, Department of Civil & Environmental Engineering, MIT, USA, 1995

Fudenberg D., J. Tirole, Game theory, MIT press, Cambridge, Mass., USA, 1991

A Semantic Information Model for Data Integration Across the Chemical Process Design Process

Andreas Wiesner, Jan Morbach, Wolfgang Marquardt

AVT-Process Systems Engineering, RWTH Aachen University, 52056 Aachen, Germany

Abstract

Information integration during the design process of chemical plants is a long-standing and not sufficiently solved problem in industrial practice to date. The major challenge identified is the capturing of the information's semantics. Within an ongoing research project an ontology-based approach for information integration in process engineering design projects is developed. This contribution sketches the semantic information model applied in the integration software, which is based on the formal ontology OntoCAPE.

Keywords: Ontology, CAPE, information modeling, information integration, design process

1. Introduction

The design of a chemical plant is characterized by a sequence of design phases, which involves different departments, disciplines, and contractors. In order to handle the design information, many software tools are used to support the project teams in the different phases, and a multitude of documents in different electronic formats are produced. Some of these software tools are domain-specific (i.e. process simulators, CAD or CAE systems, etc.) and others are generic and independent of the requirements of chemical engineering design processes (i.e. word processors, project management systems, etc.). However, the contents of these documents and data files typically are heavily intertwined and overlapping due to strong dependencies between the information generated in the different phases. Unfortunately, the information models of domain-specific software tools suffer from the lack of a well-structured, standardized information representation [1,2]. Almost all of them lack capabilities for interoperability and are thus mainly applied in a stand-alone manner. In order to cover design data over the entire design process, many information modeling activities, e.g., STEP [3], have been carried out, not only but also in the process engineering domain. ISO 15926 [4] is a prominent example. These formats are intended to be used for data exchange in order to integrate software tools during the engineering design process.

However, the most prevalent problems of information integration have not been successfully addressed by any of the current approaches: (i) consistency checking, (ii) the determination of dependencies between the contents of the documents, or (iii) the homogenization of different proprietary data formats. One reason for that seems to be an insufficient capturing of the semantics of the information modeled. Hence, we are exploring the use of more expressive modeling approaches to facilitate the capturing of the detailed semantics of the modeled information. Within an ongoing research project an ontology-based software prototype for the integration (consolidation and reconciliation) of distributed design data stemming from different design phases is developed. This contribution focuses on the prototype's semantic information model

which is based on the domain ontology OntoCAPE [5,6]. For details on the conceptual design of the software prototype and its implementation we refer to [7,8].

The remainder of the paper is organized as follows: Section 2 gives a brief overview on the typical design phases and the resulting documents. Section 3 introduces the principles applied to formally represent design aspects in OntoCAPE. In Section 4, the connection between the design aspects in OntoCAPE and the data stored in the software tools required for compliance checking is shown. Finally, Section 5 concludes the contribution.

2. Design phases and its corresponding documents

Like in any other engineering domain, the chemical plant design process can be subdivided in a number of distinct phases [9]. Typically in the first phase, the requirements for the material products to be produced by the chemical plant are fixed. Besides various business objectives (e.g. the aspects of market demand and the current and expected supply by competitors for the estimation of production capacity), the material product properties identified by the chemists are of major importance for the actual process design. These requirements form the basis for a first description of the design problem. However, the requirements cannot be completely captured and expressed by common CAE systems such that formal representations of requirements are rarely available in real-world projects. Typically, they are stored as text files, data sheets or simply presentation slides.

The design problem formulation is further refined in the conceptual design, where the major conceptual decisions on raw materials, on the chemical synthesis route, on the process structure, and even on the strategy for plant operation are taken. These conceptual considerations are refined in front-end engineering and detailed and completed during basic engineering [9,10]. At the end of basic engineering all major design data of the plant are fixed. The important documents at the end of front-end engineering are the PFD and so-called process data sheets for the most important pieces of equipment as well as a skeleton P&ID for basic engineering.

Basic engineering is followed by detail engineering, where the design data fixed during basic engineering are used to specify all pieces of equipment of the plant including all the instrumentation and control systems in full detail. The design process is completed at the end of detail engineering. The result of the design process is a set of complete specifications which are on a level of detail to allow procurement of all parts and construction of the plant on site in the subsequent plant lifecycle phases. This phase is usually documented by means of full-fledged P&ID's (which represents the major equipment items and their main dimension, but no geometrical details) and on a more detailed information level attained by isometric drawings and 3D models.

3. Representation of design aspects in OntoCAPE

OntoCAPE is a formal ontology for the domain of Computer-Aided Process Engineering (CAPE). An ontology is an explicit specification of a conceptualization, typically involving classes, their relations, and axioms for clarifying the intended semantics [11]. A formal ontology, in particular, captures the consensual knowledge of an application domain in such a way that it can be reused and shared across software systems. OntoCAPE formally represents domain knowledge about CAPE, with the intention of enabling the construction of software for different tasks such as knowledge management or plant design. Terms corresponding to classes defined in OntoCAPE are written in *italic* font in the following.

OntoCAPE is organized according to the principles of general systems theory[1] and systems engineering[2]. It introduces important systems-theoretical and physicochemical primitives such as *system*, *property*, *value*, etc., and specifies their mutual relations (e.g., a *system* is characterized by its *properties*, each of which can take numerous *values*). It also establishes the notion of so-called *aspect systems*, which represent a system considered from a particular viewpoint[3] which are used to partition complex systems into manageable parts. For a comprehensive description of *technical systems,* i.e. a system that has been developed through an engineering design process, five designated viewpoints are of major importance [1]: the *requirements* and the *function* of the *system*, as well as its *realization*, its *behavior*, and its *performance*. However, within this contribution we focus on the first three viewpoints as they are mainly related to the major design documents generated in a typical design process. For a comprehensive description we refer to [12].

First of all, the precise meaning of the *requirements* and *function* viewpoint is clarified. The ontological representation of function in design is a long-standing research issue. Various definitions of the function concept have been proposed in the literature; for a review of those, see for example [13]-[15]. For our purposes, we adopt the definition of [13], who defines function as desired behavior. According to these authors, two interpretations of the function concept must be distinguished for a *technical system*: function seen from an environment-centric viewpoint and function seen from a device-centric viewpoint. In OntoCAPE, a system's *requirements* specify the desired effect of the *technical system* on the environment. From the perspective of *requirements*, the *technical system* is viewed as a black box: Its structure and the underling physical and technical principles are not considered; only the effect on the environment is specified. However, a system's *function* describes the intended behavior of a *technical system* from a device-centric perspective. To indicate the system's *function* of a *technical system*, the conceptual design of the *technical system* must be specified in terms of the underlying physicochemical and/or technical principles. As an example, consider the design of a process unit. The system's *requirements* can be stated by describing the effect that the process unit shall exert on the processed materials (e.g., "separate dispersed particles from a liquid"). Yet to specify the system's *function*, one needs to consider the physical or technical principles based on which the desired effect is going to be achieved (e.g., decide whether the separation is realized by means of sedimentation, centrifugation, or filtration).

The third *aspect system* to be described here is the *realization* viewpoint. It reflects the physical (or virtual) constitution of the *technical system*. In case of a chemical process system, the *realization* describes the system's physical structure with respect to its geometrical and mechanical properties, e.g. comprising the equipment and machinery required for materials processing. Generally, the *realization* gives a static description of the technical system which is comparable to a technical specification, as it is typically created in an engineering design project to specify the technical system that is to be built. In this context, it is important to remember that a system realization holds only

---

[1] General systems theory is an interdisciplinary field that studies the structure and properties of systems [16].

[2] Systems engineering can be viewed as the application of engineering techniques to the engineering of systems, as well as the application of a systems approach to engineering efforts [17].

[3] The adoption of a viewpoint is a technique for complexity reduction that is widely used in systems engineering. A viewpoint is an abstraction that yields a specification of the whole system restricted to a particular set of concerns [18].

information pertaining to the system itself; information that specifies how to realize a technical system (e.g., assembly instructions) does not form part of the *realization*.

Finally, important relations and dependencies exist between the *aspect systems* of a *technical system*: In the course of a design project, the *requirements* built the cornerstones of the design project. Thus, the system's *function* as well its *realization* has to comply with them. This is usually guaranteed by the fact that the system's *requirements* are directly transformed into system's *functions*, which specify the conceptual design. Similarly, the system's *function* is detailed into the system's *realization* at the stage of basic design.

4. Realization for compliance checking for design data

Due to the lack of software vendor solutions to remedy interoperability, most of the major chemical and engineering companies have established in-house solutions for data exchange. Typically, these solutions are based on XML such that tailored converters allow the import and export of the respective design data. However, the major problem in today's engineering departments is to achieve consistency for the data set within the course of an entire design project. Regardless of the project phase, it must be ensured that the suggested design still complies with requirements as envisioned in the first phase of the design process, e.g., each refinement has to meet the constraints formulated in an earlier design phase. Not being compliant with predefined requirements often leads to unnecessary revision of the design data which usually causes considerable extra cost. Therefore, the novel software prototype, which is currently under development, aims at two goals [8]: Firstly, it allows for data exchange of XML files between tools which support object-oriented data export/ import; secondly and most importantly it allows for comprehensive consistency checks. To that end, the main emphasis is placed on compliance checking of (i) technical realization data (e.g., the interconnection of flanges with different internal diameters or process parameters exceeding the stability limits of some plant equipment) as well as of (ii) product specification data (e.g. product X must not exceed temperature T_{max} to avoid isomerisation of X to X*).

In the following, we give a simplified example on how the proposed information model can be employed to guarantee the required data consistency by considering example (ii) referring to product specification constraints. Such constraints, usually formulated at the very beginning of the design project, may initially be described, according to the principles stated in OntoCAPE: The *temperature property* characterizing *material* (chemical) X may take at most the *value* T_{max}, regardless of the *(sub)system* comprising the *material*. This requirement has to be formally defined by the classes and relations provided by OntoCAPE and stored in the knowledge base. It can then be used for consistency checks with the data generated in the following phases in the design project. The data captured in a PFD, as provided at the end of conceptual design, may be adequately described by the *aspect system* of *function*. Typically a PFD is an aggregation of blocks and streams which can be represented by special (sub)systems in OntoCAPE. Blocks may be represented by so-called *process steps*, i.e. a certain material processing procedure, and streams are depicted by so-called *process states* which represent the collection of *properties* of a certain *material* produced in the associated *process step* [19]. Accordingly, all *process steps*/*process states* referring to *material* X, which are characterized by *temperature properties,* can now be reconciled to the former formally stated *requirements*. Proceeding to the next stage of the design project, i.e. front-end engineering/basic engineering, the data are documented by process data sheets and (skeleton) P&IDs. At this point of the design project the data correspond to *plant*

items actually to be constructed. Accordingly the *realization* aspect system is employed subsequently. The realization viewpoint generally consists of the major plant items such as *equipment*, *instrumentation* and *piping* which are again special subsystems and thus conform to the principles stated before [19]. Since the equipment process data are typically determined in this design phase, each *equipment* processing *material* X can be checked against the *process steps* specification for consistency to guarantee that the range of the *temperature property* is within limits. Finally, if it comes to the stage of basic engineering/detail engineering, the equipment is further refined to a level of detail to allow construction. Also the remaining plant items, essentially instrumentation and piping, are specified. To meet the consistency constraints relating to temperature, the technical specification data of *equipment* may be reconciled against the process data and the configuration of *instrumentation* and *piping* respectively may be checked against the *requirements* alternatively. Despite the fact that the technical specification of equipment, instrumentation and piping is achieved in the same design phase, it is typically done by different disciplines and thus different engineers. Hence, the alignment of technical realization data is an additional issue which, however, can be checked with our approach as well. Due to the lack of space we refer to [7,8] for a more detailed description on that issue.

Thus, coherent information integration, in the sense of obtaining consistent data sets, can be achieved by the formulation of constraints. Engineers have to handle a variety of constraints and their exceptions in the course of a typical design project which have to be classified in categories. The formulation of constraints and their exceptions is best achieved expressing them declaratively as rules. Hence, all aforementioned consistency checks are realized via rules. These rules represent declarative knowledge in the form "if A then B", where A and B are statements about the information that is expressed by means of ontological terms. By choosing a suitable ontology language [20] and inference engine (reasoner) [21], which is capable of deductive reasoning, the rules can be executed. Considering example (ii) above such a rule may state (informal representation): **IF** X is a *chemical material* **AND** it is contained in a *(sub)system* **AND** the *material* is characterized by a *temperature property* **AND** the corresponding *value* exceeds T_{max} **THEN** the *(sub)system* is declared inconsistent.

5. Conclusions

This contribution presents an outline of the semantic information model of a novel, ontology-based approach for information integration for process engineering design data. The model is based on the formal ontology OntoCAPE, which provides the necessary principles and vocabulary to formally express information such that automatic data reconciliation by means of suitable software components (reasoners) is achievable across the phases of a design process. Therefore, we introduced the representation of certain viewpoint on the design data within a design project such that design data stemming from different phases and disciplines may be defined according to its information content. Hence, this approach is capable of successfully handle: (i) hidden dependencies between the contents of the documents applied in the design project; and (ii) compliance checking for the design data based on the categorization in so-called aspect systems. So far, the compliance checking between the viewpoints of realization and requirements took priority for the development of the software prototype. It has been tested against small to medium size amounts of real plant data and has been able to fulfill all needs with respect to technical realization as well as product property constraints. Even highly complex real world data could be inferred correctly. Thus, the

approach constitutes the semantic basis for successful information integration as envisioned by industrial practitioners.

References

[1] B. Bayer, (2003), Conceptual Information Modeling for Computer Aided Support of Chemical Process Design. Fortschritt-Berichte VDI: Reihe 3, Nr. 787. VDI-Verlag, Düsseldorf.

[2] B. Bayer and W. Marquardt, (2004), Towards integrated information models for data and documents, *Comp. Chem. Eng.*, 28 1249-1266.

[3] ISO10303, (1998), Standard for the Exchange of Product data

[4] ISO15926, (2003), Industrial Automation Systems and Integration – Integration of Life-Cycle Data for Process Plants Including Oil and Gas Production Facilities

[5] W. Marquardt, J. Morbach, A. Wiesner, A. Yang, (2009), OntoCAPE – A (Re)Usable Ontology for Chemical Process Engineering. Springer, Berlin. In preparation.

[6] J. Morbach, A. Wiesner, W. Marquardt, (2009), OntoCAPE 2.0 – a (Re)Usable Ontology for Computer-Aided Process Engineering . *Comp. Chem. Eng.*, in press.

[7] J. Morbach and W. Marquardt (2008), Ontology-based integration and management of distributed design data. In: Nagl, M., Marquardt, W. (eds.): *Collaborative and Distributed Chemical Engineering*. Springer, Berlin, pp. 647–655.

[8] A. Wiesner, J. Morbach, W. Marquardt, (2008), Semantic data integration for process engineering desgin data. In: *Proceedings of the 10th International Conference on Enterprise Information Systems – ICEIS 2008*, pp. 190-195

[9] M.S. Peters, K.D. Timmerhaus, (1991), Plant Design and Economics for Chemical Engineers. McGraw-Hill

[10] J.M.Douglas, K.D. Timmerhaus, (1988), Conceptual Design of Chemical Processes. McGraw-Hill

[11] M. Uschold and M. Grüninger, (1996), Ontologies: principles, methods and applications. *Knowl. Eng. Rev.* 11 (2), 93–136.

[12] J. Morbach, B. Bayer, A. Wiesner, A. Yang, W. Marquardt, (2009), OntoCAPE 2.0 – The Upper Level. Technical Report (LPT-2008-25), Lehrstuhl für Prozesstechnik, RWTH Aachen University.

[13] B. Chandrasekaran, J.R. Josephson, V.R. Benjamins, (1999), What are ontologies, and why do we need them? *IEEE Intell. Syst.* 14 (1), 20–26.

[14] S. Szykman, R.D. Sriram, W.C. Regli, (2001), The role of knowledge in next-generation product development systems. ASME J. *Comp. Inf. Sci. Eng.* 1 (1), 3–11.

[15] Y. Kitamura, and R. Mizoguchi, (2003), Ontology-based description of functional design knowledge and its use in a functional way server. *Expert Syst. Appl.* 24 (2), 153–166.

[16] L.v. Bertalanffy, (1968), General System Theory: Foundations, Development, Applications. Braziller, New York.

[17] B. Thomé, ed. (1993), Systems Engineering: Principles and Practice of Computer-based Systems Engineering. Wiley, New York.

[18] IEEE, (2000), IEEE Recommended Practice for Architectural Description for Software-Intensive Systems. IEEE Standard 1471-2000, Institute for Electrical and Electronics Engineering, New York.

[19] A. Wiesner, J. Morbach, B. Bayer, A. Yang, W. Marquardt, (2009), OntoCAPE 2.0 – Chemical Process Systems. Technical Report (LPT-2008-29), Lehrstuhl für Prozesstechnik, RWTH Aachen University.

[20] M. Kifer, G. Lausen, J. Wu, (1995), Logical foundations of object-oriented and frame-based languages, *Journal of the ACM* (42:4).

[21] Ontobroker, (2009), http://www.ontoprise.de/en/home/products/ontobroker/

[22] M. Nagl and W. Marquardt, eds. (2008), Collaborative and Distributed Chemical Engineering: From Understanding to Substantial Design Process Support. Springer, Berlin.

A Parallel Computing Scheme for Large-Scale Logistics Network Optimization Enhanced by Discrete Hybrid PSO

Yoshiaki Shimizu, Takatobu Miura and Masashi Ikeda

Department of Production Systems Engineering, Toyohashi University of Technology,
1-1 Hibarigaoka, tenpaku-cho,Toyohashi 441-8580, Japan

Abstract

Recently, we have concerned the strategic optimization on logistic network design problems and developed an efficient two-level solution method using a meta-heuristic method like tabu search. To cope with extremely large-scale problems, in this paper, we propose an extended algorithm for the parallel computing that utilizes the analogy between the algorithm and the master-worker configuration of PC cluster. To enhance the efficiency, we developed a modified discrete algorithm of particle swarm optimization (PSO) that can deal with binary decision variables. Then, we applied it to a parallel procedure that requires only very small overhead for the parallel computing. Finally, we confirmed that the proposed method can bring about high performance for the parallel computing that is suitable for the present goal and circumstance through numerical experiments.

Keywords: Logistic network optimization, PSO, Parallel computing algorithm, Master-worker PC cluster.

1. Introduction

Noticing the growing importance of supply chain management in manufacturing, we have engaged in a logistics network optimization using a method termed hybrid tabu search, and deployed it in various situations (Wada, Shimizu & Yoo, 2005; Shimizu, et al., 2006; Shimizu, et al., 2007). Compared with previous studies made extensively (for example, Drezner, 2002), these studies can cope with large-scale problems in a practical and flexible manner. To cope with extremely large-scale problems in a numerically effective manner, in this paper, we propose an algorithm for the parallel computing aimed at reducing the computation time and improving the optimality at the same time. For this purpose, we have given a discrete algorithm of particle swarm optimization (PSO) so as to deal with binary variables that stand for open or close of sites and introduced a mutation operation like genetic algorithm into the conventional PSO. Eventually, we will show such a population-based algorithm is particularly suitable for the parallel computing targeted at global logistics optimization. After showing the algorithm of such modified discrete PSO, we will outline the procedure of the parallel computing and discuss its properties. Finally, we provide a few numerical experiments to validate the effectiveness of the proposed method.

2. Parallel Computing for Logistic Optimization

2.1. Strategic Logistics Network Model

Let us take a logistic network composed of plant, distribution center (DC), and customer as shown in the left side of Fig.1. Then consider an optimization problem formulated

as the following mixed-integer programming problem.

$$Min \quad \sum_{i \in I}\sum_{j \in J}(T1_{ij} + C_i)\cdot e_{ij} + \sum_{j \in J}\sum_{k \in K}(T2_{jk} + H_j)\cdot f_{jk} + \sum_{j \in J}F_j \cdot x_j$$

subject to

$$\sum_{j \in J}f_{jk} = D_k \quad \forall k \in K \tag{1}$$

$$\sum_{i \in I}e_{ij} = \sum_{k \in K}f_{jk} \quad \forall i \in I \tag{2}$$

$$\sum_{j \in J}e_{ij} \leq S_i \quad \forall i \in I \tag{3}$$

$$\sum_{k \in K}f_{jk} \leq U_j \cdot x_j \quad \forall j \in J \tag{4}$$

$$x_j \in \{0,1\} \quad \forall j \in J$$

$$e_{ij}, f_{jk} \geq 0, \ \forall i \in I, \ \forall j \in J, \ \forall k \in K$$

where notations denote as follows:

C_i: production cost per unit amount at plant i
D_k: demand of customer k
F_j: fixed-charge cost for opening DC j
H_j: holding cost per unit amount at DC j
S_i: upper bound for production at plant i
$T1_{ij}$: transport cost from plant i to DC j per unit amount
$T2_{jk}$: transport cost from DC j to customer k per unit amount.
U_j: upper bound of holding capacity at DC j
e_{ij}: shipped amount from plant i to DC j
f_{jk}: shipped amount from DC j to customer k
x_j: take 1 if DC j is open, otherwise 0
I, J, K: index sets of plants, DCs and customers, respectively

Fig.1 Transformation procedure into MCF graph

The objective function is the total cost composed of transportation costs, production costs at plants, holding costs at DCs, and fixed-charge costs for opening the DCs. On the other hand, we impose the constraint on demand satisfaction for every customer, Eq.(1); input-output balance at each DC, Eq.(2); available amount of product from each plant, Eq.(3); and upper bound of holding capacity at each DC, Eq.(4). In addition, binary decision variables are introduced for selecting opening DCs and non-negative real variables for deciding transport amounts.

To solve this kind of problem, we successfully applied the hybrid tabu search. Its development relies on the fact that meta-heuristic method is amenable for solving the upper-level sub-problem that decides the available DCs and refers to a 0-1 programming problem. On the other hand, the route selection problem from plants to customers via DCs in the lower-level reduces to a linear program after binary variables are fixed. Additionally, as illustrated in the right side of Fig.1, it can be further transformed into the minimum cost flow problem by introducing a few virtual nodes and edges. Consequently, we can solve the lower-level sub-problem by the graph algorithm much faster than by the linear programming method.

Then, noticing the analogy between such a two-level algorithm and the master-worker configuration of the PC cluster (See Fig.2), we proposed a suitable framework for the parallel computing for this logistics optimization. There, the master PC engages in deciding the DC location. That corresponds to the upper-level sub-problem while each worker tries to solve the route selection sub-problem in the lower-level with DC sites appointed by the master. Generally

Fig.3 Schematic idea of single and multi-walk searches

speaking, the global optimum is achievable more efficiently by the population-base algorithm than by the single-start local search algorithm like tabu search. For this purpose, we employ the PSO after giving its modified algorithm mentioned below. It can readily realize a multi-walk search in the parallelism since each worker can entirely engage in the route selection problem as shown in Fig.3.

2.2. Modified PSO for 0-1 Programming Problem

The PSO is a meta-heuristic optimization technique developed recently (Kennedy and Eberhart, 1995, 1997) after the behavior of bird flocking or fish schooling (swarm), and known as a powerful global optimization method of real variables.

Members of swarm communicate with each other and adjust their own positions and velocities based on the information regarding good positions both of their own and the swarm. In practice, the position and the velocity are updated through the following formulas, respectively.

$$x_i(t+1) = x_i(t) + v_i(t+1), \tag{5}$$

$$v_i(t+1) = w \cdot v_i(t) + r_1 b(p_i - x_i(t)) + r_2 c(y_n - x_i(t)), (i = 1, 2, \ldots, N_p), \tag{6}$$

where t denotes a generation, N_p a swarm size, and w an inertial constant. Moreover, b and c are constants that will properly guide each member to a good position. On the other hand, r_1 and r_2 are random values in the range [0,1], p_i is the best position seen by the member i (personal best), and y_n is the global best position seen by the swarm (net best).

The algorithm is simple and outlined below.

Step 1: Set t=1. Initialize $x_i(t)$ and $v_i(t)$ randomly within the admissible range of these
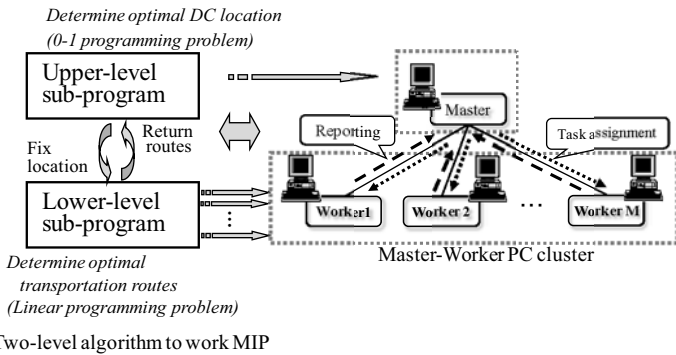
Fig.2 Analogy between two-level algorithm and Master-worker PC cluster

values, each p_i to the current position, and y_n to the position having the best fitness among the swarm.

Step 2: For each member, do the following: obtain $x_i(t+1)$ and $v_i(t+1)$ according to Eqs.(5) and (6), respectively, and evaluate the new position. If it outperforms p_i, update p_i, and if it outperforms y_n, update y_n.

Step 3: If the stopping condition is satisfied, stop. Otherwise, let $t:=t+1$, and go back to Step 2.

Now, to cope with 0-1 programming problems, we modified the original method as a discrete PSO. First, let us give a binary code of decision vector so that it can describe whether the site is open (1) or close (0) as exemplified in Fig.4. By noting that the bit-wise calculation of $p_i - x_i(t)$ or $y_n - x_i(t)$ in Eq.(6) takes anyone of (-1, 0, -1), we update the velocity at the next generation as shown in Table 1.

Table 1 Decision method of the next velocity instead of Eq.(6)

Pattern	$v_i(t+1)$	Probability
$p_i(t) - x_i(t) = y(t) - x_i(t) = v_i(t)$	$v_i(t)$	$r_a (= 0.997)$
$y(t) - x_i(t) \neq p_i(t) - x_i(t) = v_i(t)$	$v_i(t)$	$r_b (= 0.7)$
$p_i(t) - x_i(t) \neq y(t) - x_i(t) = v_i(t)$	$v_i(t)$	$r_c (= 0.6)$
$p_i(t) - x_i(t) = y(t) - x_i(t) \neq v_i(t)$	$y(t) - x_i(t)$	$r_d (= 0.5)$
$p_i(t) - x_i(t) \neq y(t) - x_i(t) \neq v_i(t)$	Select randomly	-

Apparently, these probabilities must satisfy the following relations such that: $1 > r_a > r_b > r_c > r_d > 0$. Likewise, the position is decided by the bit-wise calculation between $x_i(t)$ and $v_i(t+1)$. That is, $x_i(t+1)$ becomes 0 if the resulting value is less than 0, while 1 if greater than 1.

To enhance the performance of the algorithm, we introduced an idea to guide forcibly some inferior members to appropriate positions (we call this "warp" hereinafter). Since this warp operation is not ruled by Eqs.(5) and (6), it can be viewed as a mutation operation in the genetic algorithm. We prepared the following three methods to carry out the warp.

Warp 1: Let move the worst member to the vicinity of the net best. The vicinity is given by the flip-flop of only one locus of the net best. This presents the intensification to the search at the expense of the loss of diversification.

Warp 2: Let move the member chosen randomly at an arbitrary position. This makes increase the diversification but disturb the intensification.

Warp 3: First let choose an arbitrary member, and then randomly copy codes from net best. This is viewed as another local search around the net best.

2.3. Algorithm for Parallel Computing by Modified Discrete PSO

In what follows, we outline the tasks assigned to master and worker PCs, respectively.

1	0	1	0	0	1	⋯	1

Fig.4 Code of location (open:1, close:0)

Master PC

Step 1: Notify the number of DC site to every worker PC.

Step 2: Enter "waiting" mode

Step 3: Compare the reported personal best with the current net best. Then update if the reported personal best outperforms the current best.

Step 4: If the reported personal best is inferior to the net worst, rewrite the current net worst, and indicate the worker to carry out the warp as far as a certain condition is satisfied. Otherwise, return the current net best to the worker.

Step 5: If a certain convergence condition is satisfied, let every worker stop, and let the net best be the final solution. Otherwise, go back to Step 2.

Worker PC

Step 1: According to the notified DC number, decide the initial location.

Step 2: Solve the route selection problem. If the current search could find the personal best, update it. Otherwise, go to Step 4.

Step 3: Contract with the master and follow its indication, i.e., either warp or to receive the latest net best.

Step 4: Re-locate the DC sites based on the PSO algorithm mentioned already and go back to Step 2.

2.4. Evaluation of the Parallel Computing

There are several factors that are popularly known to affect on the performance of the parallel computing, i.e., the load imbalance between master and worker, the granularity and the frequency of communication, and the overhead for the parallelism. Regarding these factors, the proposed framework has nice properties. First, it can provide a wonderful timing that enables us to almost completely avoid the idle time due to the load imbalance. Regarding the information exchange between the master and worker, not only its amount is small but also its frequency is low due to the multi-walk implementation (The search starting from the different point is totally referred to each worker PC). Moreover, synchronization between every PC is quite unnecessary regarding information exchange. Consequently, each member is to be controlled by the different net best by the virtue of asynchronization. Without paying any particular attentions, this brings about the increase in manifoldness that is essential for the meta-heuristic algorithm. Due to these effects, we can make the overhead for the parallelism very small and improve the performance of the algorithm at the same time.

To evaluate the total performance of the parallel computing, the following two indices are commonly used in the homogeneous computing environment. When it takes $T(P)$ CPU time using P number of PCs, speedup rate is defined by $S(P) = T(1)/T(P)$, and the efficiency by $\eta = S(P)/P$. Hence, these values ideally become $S(P)=P$ and $\eta = 1$, respectively.

3. Numerical Experiments

Using up to 9 PCs with various specifications, one of which works as the master while the others as workers, we performed numerical experiments. They are run under the Debian Linux and using the library of MPICH (Pacheco, 1997). We prepared benchmark problems with the problem size such as $|I|$=30, $|J|$=300, and $|K|$=1000. When number of evaluation has reached at 8000, we stop the search. To properly instruct the poor worker regarding the warp, we gave a condition that relates to the improvement rate, i.e., $\Delta Cost / \Delta Time \le D$ where $\Delta Cost$ and $\Delta Time$ denote the difference of objective function between the solutions and of evaluation time, respectively. Moreover, D is an appropriate constant controlling the timing of the warp. Adjusting this condition, we can avoid the undesirable warps,
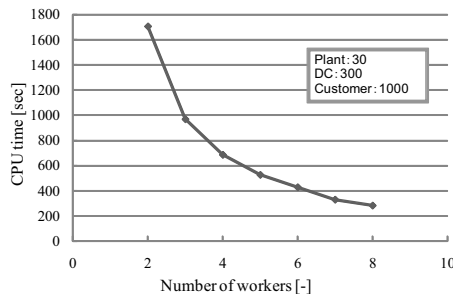
Fig.5 Effect of number of workers on speed up

i.e., at the beginning of the search and during the active updating stage. Table 2 summarizes the results (averaged over ten trials) among various warp methods. From Table 2, we know every warp method achieved the better performance than the original method without any warp, i.e., less computation time and objective value. Besides the better convergence property, the speedup rate shown based on the wall-clock time in Fig.5 also validates the effectiveness of the proposed method. Thereupon, the parallelism effect is splendid and moderately degraded with the increase in the number of workers. From these facts, we can expect to efficiently solve much larger-scale problem using the larger cluster by the proposed method.

Table 2 Comparison of performance among the warp methods

	Objective value	CPU time
Original (No warp)	33601	275.9
Warp 1	33464	223.2
Warp 2	33581	219.6
Warp 3	33569	212.7

4. Conclusions

To solve strategic optimization problems for large scale logistic network design, we have developed a practical and efficient method for the parallel computing and implemented it into a master-worker configuration. Thereupon, we are interested in the PSO and applied it after giving its modified algorithm to handle binary variables.

This is a novel approach of PSO for realizing information sharing and exchange and multi-walk in the parallel computing. Additionally, due to the analogy between the two-level solution algorithms and the master-worker configuration of PC cluster, we have shown the developed procedure can provide very nice properties regarding the reduction of the overhead and the idle time. As a side effect, the idea can increase manifoldness that is essential for the population-based meta-heuristic approach.

Through numerical experiments, the effectiveness of the proposed procedure is confirmed. Relying on these results, we can conclude that the proposed approach is promising for large-scale and complicated real world applications, targeting at global logistics of chemical industries.

References

J. Kennedy and R. C. Eberhart, 1995, Particle Swarm Optimization, Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ., 1942-1948

J. Kennedy and R. C. Eberhart, 1997, A Discrete Binary Version of the Particle Swarm Algorithm, Proceedings of World Multiconference on Systemics,Cybernetics and Information, Piscataway, NJ., 4104-4109

Z. Drezner and H. W. Hamacher (eds.), 2002, Facility Location, Springer, Berlin

P. S. Pacheco, 1997, Parallel Programming with MPI, Morgan Kaufmann Publisher

Y. Shimizu, S. Matsuda and T. Wada, 2006, A Flexible Design of Logistic Network against Uncertain Demands through Hybrid Meta-Heuristic Method, Proc. 16th ESCAPE, 2051-2056

Y. Shimizu, T. Wada and Y. Yamazaki, 2007, Logistics Optimization Using Hybrid Meta-heuristic Approach under Very Realistic Conditions, Proc. 17th ESCAPE, Bucharest, Romania

T. Wada, Y. Shimizu and J. K. Yoo, 2005, Entire Supply Chain Optimization in Terms of Hybrid in Approach, Proc. 15th ESCAPE, 1591-1596

Y. Shimizu, H.Kawamoto, 2008, An Implementation of Parallel Computing for Hierarchical Logistic Network Design Optimization Using PSO, Proc. 18th ESCAPE, 605-610

Supply Chain Optimization for Bioethanol Production System in Northern Italy: Environmentally Conscious Strategic Design

Andrea Zamboni[a], Fabrizio Bezzo[a], Nilay Shah[b]

[a]*DIPIC – Dipartimento di Principi e Impianti di Ingegneria Chimica,Università diPadova, via Marzolo 9, Padova, I-35131, Italy (andrea.zamboni@unipd.it).*
[b]*CPSE – Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ,UK*

Abstract

This work proposes a spatially explicit modeling framework developed for the strategic design of biomass-based fuel supply networks. A Mixed Integer Linear Programming (MILP) model is formulated and solved for the multi-objective optimization of the whole supply network in terms of both operating costs and greenhouse gas (GHG) emissions. The economics are assessed by means of Supply Chain Analysis (SCA) techniques. The environmental performance is evaluated in terms of greenhouse gas emissions, by adopting a Well-to-Tank (WTT) approach. The framework capabilities in steering the strategic design path are assessed through a real-world case study.

Keywords: Supply Chain Design, Multi-objective Optimization, Life Cycle Analysis.

1. Introduction

Concerns related to oil depletion as well as environmental issues like global warming have been driving a worldwide debate centered on the global energy supply question. In the transport sector, liquid biofuels have been identified as a viable and realistic alternative to fossil energy in achieving energy supply security and climate benefits in the near future. Italy has complied with the European guidelines (Directive 2003/30/EN) by setting the minimum blending of biofuels at 3% in energetic content for 2009 and 5.75% for 2010. Bioethanol is globally considered the current best viable alternatives for a short-term partial gasoline replacement. However, some doubts still persist on whether ethanol production from starchy biomass brings any effective economic and environmental benefits [1]. In particular, production costs and greenhouse gas emissions strongly depend on the geographical conditions and technological experience of each country which the production system is operating in. As a consequence, the transition path from an oil-based fuel system to a biomass-based one should be driven by specific tools capable of assessing the economic, technological and social interactions along the entire supply chain (SC).

This paper proposes a spatially explicit modeling framework developed for driving the strategic design of biofuels supply networks under economic and environmental criteria. The design task is formulated as a Mixed Integer Linear Program (MILP) that accounts for the simultaneous minimization of the SC operating costs as well as the environmental impact in terms of greenhouse gas (GHG) emissions. The economics have been assessed by means of Supply Chain Analysis (SCA) techniques, focusing on

biomass cultivation site locations, ethanol production capacity assignment and facilities location as well as transport system optimization. The environmental performance of the system has been evaluated in terms of GHG emissions, by adopting a Well-to-Tank (WTT) approach. The proposed tool has been applied to design the forthcoming Italian corn-based ethanol system. The reported results demonstrate the framework capabilities in providing valuable insights in steering the design of strategic infrastructures.

2. SC and LC analysis

The economic as well as the environmental assessment of a production system by means of SCA and LCA techniques requires a rigorous preliminary work focused on the characterization of the supply chain components as well as of the logistic nodes in terms of infrastructure availability, operating costs and environmental impact. Fig. 1 depicts the general structure of the system. A biomass-based fuel supply chain can be divided into two main substructures: the first one concerns with the fuel upstream production network and involves biomass cultivations, biomass delivery and fuel production sites; the latter is related to the downstream product distribution to the demand centers.
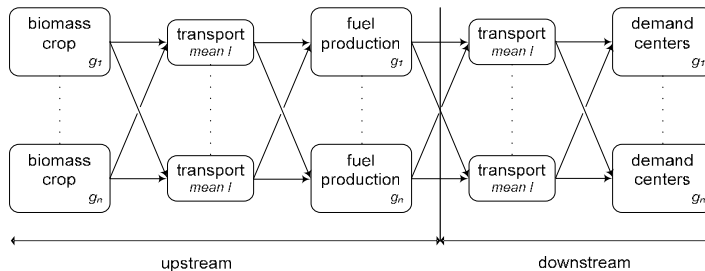
Figure 1. Bioethanol supply chain superstructure

2.1. Biomass cultivation

Considering the first generation production technology as the best solution over a short-term horizon, corn has been identified as the most convenient biomass for ethanol production in Italy. Province specific data regarding yield and land availability for corn crops have been collected from Governmental institution databases (www.istat.it; www.apat.it). Production costs have been derived from actual data (www.crpv.it) and then divided into fixed costs and yield dependent ones, so as to create a grid-dependent set of parameters. This approach has been validated by comparing the obtained results with actual production costs collected from regional databases. The same procedure has been applied to adapt specific data regarding Italian corn cultivation practice [2], in order to obtain a set of grid specific global emission factors resulting from an interactive spreadsheet based tool [3] conceived according to the IPCC guidelines.

2.2. Transport system

The Northern Italy infrastructure includes a full-scale range of transport options available for industrial purposes. The whole set has been considered defining trucks, rail, barges and ships as possible delivery means. Assuming the delivery of goods as an additional service provided by existing actors already operating within the infrastructure, transport costs have been evaluated by multiplying the freight loads by a unit transport cost [¤/(t·km)] gathered from the literature [4] and then validated comparing the resulting outcomes with actual data coming from confidential information. Global emission factors specific for the different transport means have

been taken from [5]. The availability of each transport option has been characterized through the definition of feasibility constraints on transport means suitability as well as through a tortuousity factor accounting for the different routes that each specific transport means has to go through.

2.3. Ethanol production

The dry grind process has been considered in characterizing the production facilities. Ethanol production costs are sensitive to plant capacity due to the economy of scale effect on capital and operating costs. This important issue has been taken into account considering four different discrete capacity ranges. For each range specific production and capital costs have been calculated by means of a purpose-designed financial model [6]. Also in this case, the global emission factors have been calculated as in § 2.1.

2.4. Cost allocation and emission credits

In calculating the operating costs as well as the environmental impact related to the entire supply chain operation, cost allocation and emission credits have been taken into account. Therefore, in evaluating the operating costs a quote for DDGS has been detracted considering a 20% allocation factor deduced from literature data [7]. According to [8], no credits have been assigned for land use. Hence, GHG emission credits have been only associated to products displacement by using DDGS as a valuable product for animal feed substitution.

2.5. Demand centers

Internal gasoline depots have been assumed as demand centers for bioethanol. Data about provincial gasoline demand perspectives for 2010 as well as internal depots location and maximum distribution capacity have been collected from Governmental databases (dgerm.sviluppoeconomico.gov.it). Given the demand driven nature of the optimization problem considered, demand assignment to depots must be solved as a secondary distribution problem before the supply chain optimization. The distribution problem has been based on the modeling approach commonly applied in the optimization of fuel distribution systems [9].

3. Optimization framework

Strategic decisions in designing a biofuel production network deal with geographical location of biomass cultivation sites, logistic definition of transport system and location as well as capacity assignment of production facilities. Therefore the key variables are:
1. biomass production in each cell;
2. distribution processes for biomass to arrive at the production facility;
3. location and capacity of production facilities;
4. distribution processes for fuel to be sent to blending terminals.

The core of the modeling framework is based on the MILP approaches commonly adopted in optimizing the strategic design of multi-echelons supply networks for renewable fuels: the work of [10] has been taken as standard reference in defining the superstructure of the problem; whilst concerning with the environmental issue the approach of [11] has been adapted into a more flexible mathematical formulation devised to avoid excessive efforts in both calculation and data collection.

Eq. 1 shows the economic objective function representing the SC operating costs (*TDC*):

$$\min TDC = \frac{FCC}{a} CCF + PC + TC \tag{1}$$

where the facilities capital costs (*FCC*) are annualized through a capital charge factor (*CCF*) and divided by the SC operating period (*a*); the additional terms are the production costs (*PC*), accounting for both biomass and fuel production, and the transport costs (*TC*).

The environmental objective function accounting for the GHG emissions in the supply chain (*TI*) results from the sum of the life cycle stage contributions as shown by Eq. 2:

$$\min \ TI \ = \ \sum_p f_p F_p \tag{2}$$

where F_p is the reference flow for each life cycle stage *p* and f_p is the global emission factor, representing the overall GHG emitted at stage *p* per unit of reference flow. The global emission factor (f_p) has been obtained by grouping the emission coefficient of each substance, together with the related damage factor based on the concept of 100-year global warming potentials as specified by the IPCC. All the cost variables as well as the stage-specific reference flows depend on design variables related to the ethanol and biomass production (P^T_{ig}), the product demands (D^T_{ig}) and the mass flows between grids ($Q_{ilgg'}$). The SC behavior is then captured through logical constraints and mass balances. Eq. 3 and Eq. 4 are demonstrative examples of this feature:

$$P^T_{ig} = D^T_{ig} + \sum_{l,g'} (Q_{ilgg'} - Q_{ilg'g}) \qquad \forall i,g \tag{3}$$

$$Q^{\min}_{il} X_{ilgg'} \leq Q_{ilgg'} \leq Q^{\max}_{il} X_{ilgg'} \qquad \forall i,l,g,g' \tag{4}$$

where *i* identifies the product (ethanol or corn), *l* the transport option and $X_{ilgg'}$ is the binary decision variable that is assigned the value 1 if the transportation of product *i* by means of *l* is allowed from *g* to *g'*, or 0 otherwise.

4. Case study

The modeling framework has been implemented to assess the 2010 demand scenario derived from the ethanol market penetration imposed by the current Italian regulations. Solving the two objective functions problem by adopting a multi-parametric approach [12] results in the trade-off set of Pareto optimal solutions reported in Fig. 2.
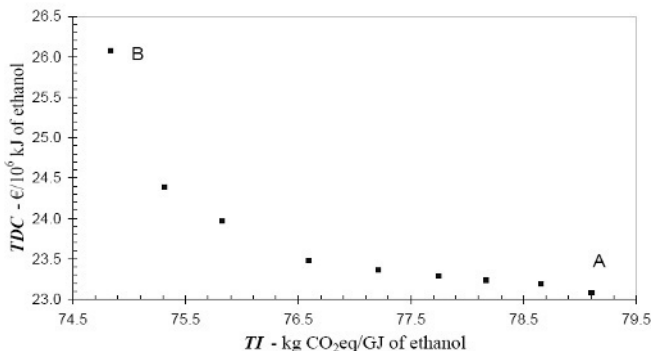
Figure 2. Multi-objective optimization: Pareto curve.

The shape of the curve reveals the expected conflict existing between environmental and economic performance. Indeed, in reaching the best optimum in terms of operating cots (point A as reported in Fig. 2, corresponding to a marginal operating costs of 23.1 ¤/10^6 kJ_ethanol) the environmental impact turn out to be 79.10 kg CO_2-eq/GJ_ethanol,

corresponding to a GHG reduction of about 8% compared to gasoline. This is not enough to meet the latest EU standards requiring biofuels to have a minimum of 35% of GHG emissions saving for them to be counted towards the target. However, even minimizing the impact (point B of Fig. 2) the resulting GHG emissions are anyhow too high to meet the minimum requirements, albeit the substantial economic effort required to reach the target: reducing the marginal impact down to 74.83 kg CO_2-eq/$GJ_{ethanol}$ (equal to 13% of GHG reduction) results in an increase of the overall operating costs up to 26.1 ¤/10^6 $kJ_{ethanol}$.

Further remarks also concerning with the market penetration of bioethanol can be based on the assessment of the performance indicators coming from the model implementation. For instance, promoting the maximum GHG mitigation in establishing the bioethanol industry would entail the need for more subsidies to fill the gap between the environmentally optimum costs performance and the economic ones (roughly amounting to 3 ¤/10^6 $kJ_{ethanol}$, about 0.1 ¤/$L_{ethanol}$). However, the corresponding environmental benefit is still not enough to satisfy with the EU standards. Therefore, in the particular case of the Italian corn-based ethanol production a well-advised strategy would address the design process under economic criteria. In this way a double benefit would be achieved: firstly the market penetration of bioethanol would be eased, so as to open the way to the oncoming second generation system; secondly, the resulting financial saving entails a bigger founds availability that might be assigned to support other energy sectors characterized by more effective environmental performance.
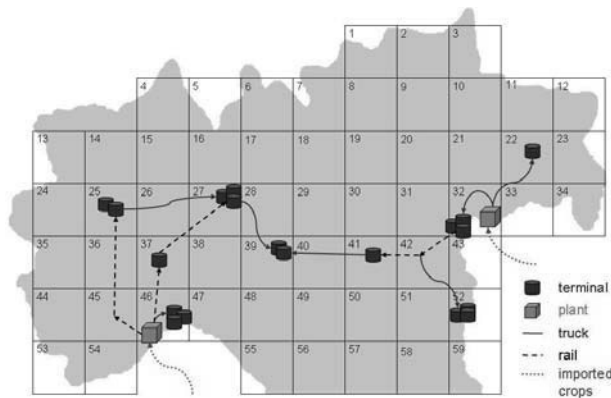
Figure 3. Costs optimization: supply network configuration.

Finally, an illustrative example of the network configuration corresponding to the costs-optimal solution is shown in Fig. 3. The graphical representation reveals important issues that both the environmental and the economic performance are dealing with. The biomass needs are met by importing corn from Eastern European Countries. Corn is directly shipped to the two production plants of the maximum capacity (about 250 kt/y) and located within the industrial areas close to the main ports of Venice and Genoa. This configuration allows the best economic performance in terms of both biomass supply costs, due to the lower price of the imported corn, and of ethanol production costs, positively affected by the scale factor. However, the longer delivery distance along with sub-optimal cultivation practice, that is often characterizing corn production in the Eastern area, are those factors that negatively affect the environmental performance. Indeed, moving along the curve shown in Fig. 2 from point A to point B, we assist to a gradual transition toward a network configuration characterized by a more decentralized fuel production system (four plant of the smaller capacity) and by

domestic biomass supply: this ensures better environmental performance in terms of goods distribution and corn production impact in spite of a drastic worsening of the system economics due to the negative scale factor as well as to the unprofitable biomass supply conditions.

A possible alternative in meeting the fuel demand would be the establishment of production plants abroad. However, this option has not been considered here mainly because it would not comply with the national policy (common in most EU countries) aiming at increasing national security in energy production.

5. Final remarks

A spatially explicit modeling framework for the strategic design of biofuel systems has been developed. The aim of the study was to build a general modeling tool that might be helpful to steer an economic and environmentally conscious design for biofuels supply chains. The optimization model applied in assessing the forthcoming Italian corn-based ethanol supply chain can be used to steer the strategic policy. The results coming from the case study application outline how the corn-based bioethanol supply network design should be based on economic criteria rather than environmental ones. This would grant an easier market penetration, as a direct consequence of the production costs minimization, and would require less governmental subsidies that thus could be deployed in more environmentally effective sectors.

6. Acknowledgement

References

[1] C. B. Granda, L. Zhu, M. T. Holtzapple, 2007, Sustainable liquid biofuels and their environmental impact. Environmental Progress*, 26,* 233–250.
[2] C. Grignani, L. Zavattaro, 2000, A survey on actual agricultural practices and their effects on the mineral nitrogen concentration of the soil solution, Europ. J. Agronomy, 12, 251–268.
[3] G. Brown, J. Wood, A. Estrin, 2005, Bioethanol greenhouse gas calculator: Users' guide, HGCA.
[4] L. Buxton, 2008, All Aboard, Biofuel International, 1.
[5] DEFRA, 2008, Guidelines to Defra's GHG conversion factors: Methodology paper for transpor emission factors.
[6] G. Franceschin, A. Zamboni, F. Bezzo, A. Bertucco, 2008, Ethanol from corn: a technical and economical assessment based on different scenarios, Chem. Eng. Research and Design, 86, 488–498.
[7] Hammerschlag, R. (2006). Ethanol's energy return on investment: a Survey of the literature 1990–present, Environ. Sci. Technol*., 40, 1744–1750.
[8] M. A. Delucchi, 2006, Lifecycle analyses of biofuel, Draft Report, University of California, Davis.
[9] M. T. Kong, 2002, Downstream oil products supply chain optimisation. PhD Thesis, University of London.
[10] A. Almansoori, N. Shah, 2005, Design and operation of a future hydrogen supply chain: Snapshot model, Chem. Eng. Research and Design, 84, 423–438.
[11] A. Hugo, E. N. Pistikopoulos, 2005, Environmentally conscious long-range planning and design of supply chain networks, J. Cleaner Production, 13, 1471– 1491.
[12] V. Dua, E. N. Pistikopoulos,2000, An algorithm for the solution of multi-parametric mixed integer linear programming problems. Ann. Oper. Res., 99, 123–139.

2043

An Effective Decomposition Approach for Solving Large Supply Chain Oriented Pick-up and Delivery Problems

Rodolfo Dondo, Carlos A. Méndez, Jaime Cerdá

INTEC (UNL - CONICET) - Güemes 3450 - 3000 Santa Fe - ARGENTINA

Abstract

Pick-up and delivery problems (PDP) are receiving a growing attention in process systems engineering due to its close relationship with major supply chain issues. Its aim is to discover the best routes/schedules for a vehicles fleet fulfilling a number of transportation requests at "minimum cost". In the conventional PDP, each request defines the shipping of a given load from a specified pickup site to a given customer. However, in order to account for a wider range of logistics problems, the so-called supply-chain oriented PDP (SC-PDP) problem has been defined as a three-tier network of interconnected factories, warehouses and customers. Multiple products are to be efficiently delivered through this network in order to meet a set of given demands. The selected vehicle routes/schedules must satisfy capacity and timing constraints while minimizing transportation costs. The pickup points for each demand are decision variables rather than problem specifications and several commodities can be transported between sites. Moreover, every customer can be visited several times. The general SC-PDP has been represented as an MILP formulation that is able to address moderate size instances. In order to efficiently address large-scale SC-PDP problems, a decomposition method based on a column generation (CG) procedure is introduced in this work. In contrast to traditional CG approaches lying on dynamic-programming-procedures as route generators, an MILP formulation is here proposed to implicitly create the set of feasible routes/ schedules at the slave level of the method.

Keywords: supply-chain, pick-up and delivery, logistics, columns generation.

1. Introduction

A typical supply chain covers the procurement of raw materials from suppliers and their shipping to one or more factories, the conversion of such inputs into intermediate and final products, their shipping to warehouses or depots for intermediate storage and the delivery of products to retailers and customers (Simchi-Levi et al., 2004). To this end, the supply chain management field and the emerging area of enterprise wide optimization have focused the attention on developing tools to efficiently coordinate factories, warehouses and customers so that product requirements are all satisfied with the specified service level at the minimum system cost. The pickup and delivery problem (PDP), widely studied in the literature, involves orders matching the pick-up of a given load at one or more locations and the subsequent delivery to one/several locations. Despite the importance of this problem, it is clear that some new features must be added to the standard PDP to obtain a better representation of realistic supply chain networks. Some of the new features to be considered are: (1) each request may involve the shipment of specific quantities of commodities to one or more destinations; (2) several commodities may be transported on the same vehicle; (3) several alternatives

sources for each commodity may be available; (4) a shipment of a given commodity from a single source may have several destinations; (5) each vehicle may be operated on more than a single route if the total time spent on these routes is less than the specified maximum service time; (6) the problem may involve multiple depots with known inventories of commodities; (7) depots may arise as intermediate stops on the vehicles routes. To face these new features, Méndez et al. [2008] proposed a variation of the PDP which involves a set of facilities from which multiple products are to be efficiently delivered to many consumers in order to meet some demands while satisfying capacity and timing constraints. This problem, named as supply-chain-PDP (SC-PDP), was addressed through a new MILP formulation that defines the choice of pickup nodes for each product demand as a set of additional decisions to be made based on the geographical locations of product demands and inventories. The higher flexibility in the routes generation process allows to find more efficient operational strategies to manage complex multi-site distribution systems. However, in addition to the development of rigorous optimization methods for logistics problems, a lot of effort of the research community has been directed to the development of heuristic procedures capable of dealing with the inherent complexity of large-scale problems. Desrochers et al. [1992] presented a new technique widely known as the *Dantzig-Wolfe decomposition + column generation* approach that remains as the most efficient optimization technique for large routing problems. Although the original SC-PDP model can be directly used to find the optimal solution for instances of moderate size, this work embeds a columns generation procedure into the model in such a way that larger examples can be solved. It is worth noting that most Dantzig-Wolfe decomposition + column generation algorithms are based on the use of dynamic-programming-label-setting algorithms for generating routes. On the contrary, the proposed procedure generates fewer columns by using a MILP-based routes-generator that arises as the adaptation to the CG structure of the model proposed by Méndez et al. [2008].

2. Problem statement

Consider a three-tier network represented by a graph G[$\{I^+, I, B\}; A$], where $I^+ = \{i_1, i_2, ..., i_n\}$ denotes the set of pick-up nodes (factories and warehouses), $I = \{j_1, j_2, ..., j_{n'}\}$ is the set of delivery nodes (customers), $B = \{b_1, b_2, ..., b_l\}$ represents the set of vehicle bases, and $A = \{a_{ij} / i, j \in I^+ \cup I \cup P \}$ defines the net of minimum cost arcs among nodes. Several nodes $i \in (I^+ \cup I)$ may be related to the same geographical location if multiple visits may be considered. A distance-based traveling-cost matrix $C = \{c_{ij}\}$ and a travel-time matrix $\Gamma = \{t_{ij}\}$ are associated to the net A. The service times on pickup/delivery nodes $i \in (I^+ \cup I)$ are denoted by st_i. To fulfill the pick-up and delivery tasks, a set of vehicles $V = \{v_1, v_2, ..., v_m\}$ is available. The solution to the problem will provide a number of sequences of arcs, commonly called *routes*, such that: (1) each vehicle starts and ends the trip at a vehicle base b; (2) for each commodity type, all the customer demands are exactly satisfied through one or multiple visits; (3) the total quantity of products provided by pick-up nodes must not exceed the initial inventory; (4) each node $i \in (I^+ \cup I)$ is at most assigned to a single route; (5) the load transported by a vehicle must never exceed both the weight q_v and the volumetric qv_v restrictions; (6) a pick-up or delivery node $i \in (I^+ \cup I)$ must be serviced within its time-window $[a_i, b_i]$; (7) the duration of the trip for any vehicle v must be shorter than its maximum allowed routing time t^{max}. All these features must be simultaneously considered while minimizing the total cost of providing pickup or delivery services to every node $i \in (I^+ \cup I)$.

3. The column generation-based approach for the SC-PDP

The SC-PDP can be formulated as a set partitioning problem (SPP) as follows: let $P = \{\rho_1, \rho_2, ..., \rho_R\}$ be the set of *all* feasible routes. Let c_ρ be the cost of route ρ and let y_ρ be a binary variable denoting that route ρ is included in the optimal solution if y_ρ is equal to 1. Then, the objective is to select the set of minimum-cost feasible routes such that each request is fulfilled. So, the optimization problem can be formulated as follows:

$$Min \quad \sum_{\rho \in P} c_\rho y_\rho \tag{1}$$

$$\sum_{\rho \in P} a_{i\rho} y_\rho \geq 1 \qquad \forall i \in \left(I^+ \cup I^- \right) \tag{2}$$

$$y_\rho = \{0,1\}$$

The parameter $a_{i\rho}$ indicates that node i belongs to route ρ if $a_{i\rho} = 1$. Otherwise, $a_{i\rho} = 0$. Since feasible routes may easily run into billions, it is not possible to realistically generate all routes (here also called *columns*), the CG approach handles this complexity by *implicitly* considering all routes through the solution of the linear relaxation of the SPP. Then, a portion of all feasible routes is enumerated and the problem linear relaxation is solved just considering this partial set. The solution to this problem is used to determine if there are routes not included in the partial set that can reduce the objective function value. Using the value of the optimal dual variables with respect to the partial routes set, new routes are generated and the linear relaxation is solved again. This procedure continues until the optimal solution to the linear problem cannot be improved with the addition of a new route. For the master problem of the SC-PDP, this basic SPP, widely used in many routing problems, must also consider inventory constraints for each product as well as the number of available vehicles. These new features leads to a new **master problem,** where the dual-variable-values are obtained by enumerating a feasible solution comprising the partial set of routes $P' = \{\rho_1, \rho_2, ...\rho_{P'}\}$. This formulation, named as the *reduced master problem* (RMP), is given below.

$$Min \quad \sum_{\rho \in P'} c_\rho y_\rho \tag{3}$$

$$\sum_{\rho \in P'} a_{i\rho} y_\rho \geq 1 \qquad \forall i \in \left(I^+ \cup I^- \right) \tag{4}$$

$$\sum_{\rho \in P'} a_{i\rho} \alpha_{i\rho} y_\rho \leq l_{ip} \qquad \forall i \in I^+, p \in P \tag{5}$$

$$\sum_{\rho \in P'} y_\rho \leq n \tag{6}$$

$$0 \leq y_\rho \leq 1$$

where l_{ip} is the inventory of product p on supply site $i \in I^+$ and n is the number of available vehicles. Now, let us assume that $\overline{y}$ is the optimal solution to the RMP and let $\Pi = [\pi_c, \pi_l, \pi_n]$ be the corresponding vectors of optimal dual variables values for constraints (4), (5) and (6), respectively. Then, if we can find a new route minimizing the quantity $(c_\rho - \Sigma_{i \in I} a_{i\rho} \pi_c^i - \Sigma_{i \in I^+} \Sigma_{p \in P} a_{i\rho} \alpha_{i\rho} \pi_l^{ip} - \pi_n)$ and, this quantity is negative, a cheaper route has been found. Consequently, $\overline{y}$ and Π are not optimal for the RMP. The column just found is added to the partial set and the problem is solved again. The procedure iterates until no columns with negative reduced costs are found. Finally, the

integer RMP can be solved for finding the best set of routes. Although in some cases the solution generated may not be the global optimal solution to the SC-PDP, it is generally very close. To find the optimal one, the procedure should be embedded into a branch & bound algorithm because some routes that are not generated when solving the relaxed RMP may be needed to solve the integer one.

The feasible routes are generated by solving the **slave problem** introduced below. The objective is to find a route ρ minimizing the quantity stated by eq. (7) subject to constraints (8) to (18). So, the *route-generator* problem can be formulated as follows:

$$Min \left[CV - \left(\sum_{i \in I} \pi_c^i Y_i + \sum_{i \in I^+} \pi_l^i Y_i + \pi_n \right) \right] \tag{7}$$

Cost-based constraints: Eqs. (8) and (9) compute the total cost for the route generated. While eqs. (8) are focused on the start and end of the trip, eqs. (9) accumulate the cost of the intermediate trips along the route. So, if nodes i and j are allocated to the route ($Y_i = Y_j = 1$), the ordering of both nodes is determined by the value of the sequencing variable S_{ij}. If location i is visited before j ($S_{ij} = 1$), according constraints (9.a), the travel cost up to location j (C_j) must be larger than C_i by at least c_{ij}. In case node j is visited earlier, ($S_{ij} = 0$), the reverse statement holds and constraint (9.b) becomes active. If one or both nodes are not allocated to the tour, eqs. (9.a)-(9.b) become redundant. M_C is an upper bound for variables C_i and CV.

$$\left. \begin{cases} C_i \geq c_{bi} \\ CV \geq C_i + c_{ib} - M_C(1-Y_i) \end{cases} \right\} \qquad \forall i \in I^+ \cup I^- \qquad \begin{array}{l} (8.a) \\ (8.b) \end{array}$$

$$\left. \begin{cases} C_j \geq C_i + c_{ij} - M_C(1-S_{ij}) - M_C(2-Y_i-Y_j) \\ C_i \geq C_j + c_{ij} - M_C S_{ij} - M_C(2-Y_i-Y_j) \end{cases} \right\} \qquad \forall i,j \in I^+ \cup I^- : i < j \qquad \begin{array}{l} (9.a) \\ (9.b) \end{array}$$

Time-based constraints: Eqs. (10) and (11) define visiting-time constraints that are similar to eqs. (8) and (9) but apply to the time dimension. M_T is an upper bound for variables T_i and TV. Eq. (12) forces the service time on any node $i \in (I^- \cup I^+)$ to start at a time T_i bounded by the interval $[a_i, b_i]$. Also, eq. (13) defines that the routing time TV must be lower than t^{max}.

$$\left. \begin{cases} T_i \geq t_{bi} \\ TV \geq T_i + st_i + t_{ib} - M_T(1-Y_i) \end{cases} \right\} \qquad \forall i \in I^+ \cup I^- \qquad \begin{array}{l} (10.a) \\ (10.b) \end{array}$$

$$\left. \begin{cases} T_j \geq T_i + st_i + t_{ij} - M_T(1-S_{ij}) - M_T(2-Y_i-Y_j) \\ T_i \geq T_j + st_j + t_{ji} - M_T S_{ij} - M_T(2-Y_i-Y_j) \end{cases} \right\} \qquad \forall i,j \in I^+ \cup I^- : i < j \qquad \begin{array}{l} (11.a) \\ (11.b) \end{array}$$

$$a_i \leq T_i \leq b_i \quad \forall i \in I \qquad (12) \qquad TV \leq t^{max} \qquad (13)$$

Cargo constraints: Eq. (14.a) states that the cargo of product p (α_{ip}) to be picked-up from any customer $i \in I^+$ must be smaller than the inventory l_{ip}, while the cargo β_{ip} to be delivered to this client must be zero. Conversely, the cargo β_{ip} to be delivered to any customer $i \in I^-$ must be l_{ip} while the cargo α_{ip} to be picked-up from this customer must be zero (eq. 14.b).

$$\left. \begin{cases} \alpha_{ip} \leq Y_i l_{ip} \\ \beta_{ip} = 0 \end{cases} \right\} \quad \forall i \in I^+, p \in P \qquad (14.a) \qquad \left. \begin{cases} \alpha_{ip} = 0 \\ \beta_{ip} = Y_i l_{ip} \end{cases} \right\} \quad \forall i \in I^-, p \in P \qquad (14.b)$$

The cargo of product p carried by the visiting vehicle up to the node $i \in (I^+ \cup I^-)$ is computed as the difference ($L_i - U_i$) between variables L_{ip} (total cargo of p loaded by the visiting vehicle up to the node i) and U_{ip} (total cargo of p unloaded by the visiting vehicle up to the node i). So, eq. (15) states that the current load of p up to the node $i \in (I^- \cup I^+)$, must be larger than zero. In turn, eqs. (16) state that the sum of these

differences must be smaller than the volumetric and weight vehicle capacity. Constraints (17) set the accumulated loaded and unloaded cargo of product p in a way similar to eqs. (9). Eq. (18.a) states that the load of product p available after visiting node i (L_i) must be larger than the quantity α_{ip} to be picked-up from the node and smaller than the quantity of goods collected on the generated tour. Constraint (18.b) is similar to (18.a) but for the cargo unloaded after servicing node i (U_i). M_L is an upper bound for L_i and U_i.

$$L_{ip} - U_{ip} \geq 0 \qquad \forall i \in I^+ \cup I^-, p \in P \qquad (15)$$

$$\left. \begin{array}{l} \sum_{p \in P} w_p \left(L_{ip} - U_{ip} \right) \leq q_v \\ \sum_{p \in P} v_p \left(L_{ip} - U_{ip} \right) \leq q v_v \end{array} \right\} \qquad \forall i \in I^+ \cup I^- \qquad \begin{array}{l} (16.a) \\ (16.b) \end{array}$$

$$\left. \begin{array}{l} L_{jp} \geq L_{ip} + \alpha_{jp} - M_L \left(1 - S_{ij} \right) - M_L \left(2 - Y_i - Y_j \right) \\ U_{jp} \geq U_{ip} + \beta_{jp} - M_L \left(1 - S_{ij} \right) + M_L \left(2 - Y_i - Y_j \right) \\ L_{ip} \geq L_{jp} + \alpha_{ip} - M_L S_{ij} - M_L \left(2 - Y_i - Y_j \right) \\ U_{ip} \geq U_{jp} + \beta_{ip} - M_L S_{ij} + M_L \left(2 - Y_i - Y_j \right) \end{array} \right\} \quad \forall i, j \in I^+ \cup I^-, p \in P : i < j \quad \begin{array}{l} (17.a) \\ (17.b) \\ \\ (17.c) \\ (17.d) \end{array}$$

$$\left. \begin{array}{l} \alpha_{ip} \leq L_{ip} \leq \sum_{j \in I^+} \alpha_{jp} \\ \beta_{ip} \leq U_{ip} \leq \sum_{j \in I^-} \beta_{jp} \end{array} \right\} \qquad \forall i \in I^+ \cup I^- \qquad \begin{array}{l} (18.a) \\ (18.b) \end{array}$$

The relaxation of the RMP is not necessarily integer and it may exist a column that would price favorably but it is not present in the columns pool. Consequently, in order to find the optimal solution we must generate columns after branching. Ryan and Foster [1981] proposed a rule that is equivalent to branch in node-to-tour assignment relationships. Following this rule, we just branch on variables Y_i. Rather than adding explicitly the branching constraints to the master problem, the infeasible columns are eliminated from the pool and branching constraints are enforced at the slave level. Summarizing, given an initial feasible solution, the series of RMP and slave problems is solved until no new columns can be found. Then, if the solution of the integer RMP (or global upper bound GUB) is higher than the solution of the relaxed RMP (or global lower bound GLB), we must branch on variables Y_i to generate the missing routes. At each tree-node, the mechanism is repeated and the bounds are compared. So, if the local lower bound is higher than the GUB, the node is fathomed; otherwise it is divided into two child-nodes that are included in the database of unsolved subspaces. Afterwards, the next subspace is fetched from the database until the base is empty. In such a case, the algorithm shows the solution and terminates. The column generation procedure is run at all nodes of the branch & bound tree and accelerating tricks as heuristic generation are used whenever it is possible.

4. An illustrative example

The proposed decomposition approach was developed with ILOG OPL Studio 3.7. A large supply chain distribution problem was solved in a 2 Ghz Pentium IV PC. The example comprises two factories (▲) located in Madrid and Barcelona from where two types of final goods (P1 and P2) are to be delivered to the 45 main Spanish cities (■). In addition, a certain quantity of auto-parts (P3) is delivered to a warehouse located in Bilbao, from where they must be transported to both factories for its posterior use. The factories host a number of identical vehicles with q_v = 15tn and qv_v = 45 m^3 that can perform pick-up/delivery tasks on customer-nodes during specified time-windows. The problem-data is summarized on Table 1. Furthermore, the average trucks travel-speed is

70 km/h, t^{max} = 120 h, service times on pick-up nodes are constants (8 h.) and unload times at delivery nodes are modeled as the sum of a constant (1 h.) and a time proportional to the quantity to deliver ($\Sigma_{p \in P}$ 0.005 P_i h; $i \in \Gamma$). The solution, detailed in Figure 1 provides the detailed routes for all vehicles. The solution, found in just 1316.4 s of CPU time, involves the use of 11 vehicles for fulfilling 12 trips with a total travelled distance of 13007 km.

Table 1. Problem data

Site	Albacete;Alicante;Almeria;Avila;Badajoz;Barcelona;Bilbao;Burgos;Caceres;Cadiz;Castellon;CReal; Cordoba;Coruna;Cuenca;Gerona;Granada;Guadalajara;Huelva;Huesca;Jaen;Leon;Lerida;Logrono;L ugo;Madrid;Malaga;Murcia;Orense;Oviedo;Palencia;Pamplona;Pontevedra;Salamanca;SSebastian;S antander;Segovia;Sevilla;Soria;Tarragona;Teruel;Toledo;Valencia;Valladolid;Vitoria;Zamora;Zarag oza
a_i	32;32;80;8;56;0;60;80;64;56;104;104;80;88;56;16;88;104;80;88;32;88;32;104;80;0;88;64;80;88;104; 56;80;32;64;88;104;88;104;88;104;8;32;8;56;64;88;80;32
b_i	40;40;88;24;64;120;68;88;72;64;112;112;88;96;64;24;96;112;88;96;40;96;40;112;88;120;96;72;88;9 6;112;64;88;40;72;96;112;96;112;16;40;16;64;72;96;88;40
$P1_i$	0;-600;-200;-300;-50;5000;-400;-200;-50;0;-300;-100;-200;-1000;-200;-400;-100;-400;-150;-200;0;- 200;-300;-500;-100;5000;-400;0;-300;0;-50;-400;-200;-800;-400;-700;-100;-600;-300;-200;-300;- 1200;-400;-200;-100;0;-700
$P2_i$	-500;0;-250;-100;0;2500;0;0;-50;-100;-200;-200;-200;-800;0;0;-250;0;-50;-200;-150;-700;0;0;- 250;2500;0;-400;-150;-600;-50;-400;-500;0;-200;-700;-50;-400;-100;-700;0;-600;-400;-500;-150;- 300;0
$P3_i$	;0;0;0;0;0;600;1000;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;400;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0;0

Figure 1. The detailed best solution

Barcelona-Gerona: 201 km/20.4 h
Barcelona-Tarragona-Lérida: 346 km/36.7 h
Madrid-Santander Madrid-Toledo: 930 km/101.6 h
Madrid-Albacete-Cuenca-Valencia-Castellón: 1116 km/113.45 h
Madrid-Lugo-Coruña: 1219 km/106.7 h
Madrid-Valladolid-León-Palencia-Segovia: 709 km/110.7 h
Madrid-Jaén-Cádiz-Córdoba-Huelva-Sevilla-CReal:1830 km/109.2 h
Madrid-Alicante-Murcia-Almería-Málaga- Granada: 1499 km/101.8 h
Madrid-Badajoz-Cáceres-Orense-Pontevedra-Oviedo: 1990 km/105.2 h

Madrid-Teruel-Zaragoza-Pamplona-Bilbao-SSebastian-Barcelona-Huesca-Guadalajara: 2138 km/118.8 h
Madrid-Ávila-Salamanca—Zamora-Burgos-Vitoria-Logroño-Soria: 1029 km/115.3 h

5. Conclusions

This work has introduced an effective optimizing decomposition approach lying on a column generation algorithm that utilizes an MILP formulation as routes generator. The proposed method is able to deal with realistic transportation problems which aim at minimizing distribution costs while keeping a high level of efficiency in complex three-tier networks comprising factories, warehouses and customers. The applicability of the approach has been proven by solving a large-scale example comprising two factories, one warehouse and forty five customers with very low computational effort.

References

Méndez C., R Dondo and J.Cerdá. Proceedings FOCAPO-2008 (2008).
Desrochers M., J. Desrosiers and M. Solomon. Op. Res. 40, 642-354 (1992)
Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E., Man. the S.Chain, McGraw-Hill (2004).
Ryan D. and B. Foster. Wren, A., (Ed.), Comp. Sch. of Public Transp. Urban Passenger.
Vehic. and Crew Sch., North-Holland, Amsterdam, pp. 269–280 (1981).

A Novel Approach to Policy Design Using Process Design Principles

Araz Taeihagh, René Bañares-Alcántara and Zun Wang

Department of Engineering Science, Oxford University, Parks Road, Oxford OX1 3PJ, UK

Abstract

Policy design, as it is currently practiced, relies on the manual formulation of policies. This fact, combined with an increase in the complexity of the systems under study, results in a large portion of the possible choices for action being left unexplored. The introduction of a systematic approach for exploring alternative policies using a computational methodology has the potential to help in the decomposition of complex problems into subproblems with more manageable size, and can accelerate and improve the effectiveness of the policy-making process. A new approach to facilitate policy design is proposed; it uses principles widely used in chemical engineering design combined with mathematical and artificial intelligence techniques. In particular, our ideas are based on the Hierarchical Design Method for the conceptual design of chemical processes (Douglas, 1988) and use of techniques such as network analysis and agent-based modelling. A six-step framework has been proposed and is being implemented as a prototype decision support system (DSS). An agent-based modelling approach, due to the opportunistic nature of agents is used as the programming paradigm in the design of policies. The DSS creates a network of policy measures and extracts information from the network during the development of policy alternatives. A set of agents then combine policy measures and goals within a tree structure at different hierarchical levels during the creation of the overall policy. The results from the research are a fresh contribution to the methodological development of policies using a multidisciplinary approach and have the potential to accelerate the design of policies and improve their success rate.

Keywords: Decision support systems, process design, policy design, transport emission reduction.

1. Introduction

As our socio-economical systems are increasingly becoming more complex, solving major national and international problems are becoming more challenging. Such complexities arise from increased number of stakeholders and the consequences of phenomena such as globalization, result in energy and environmental constrains (both intentional, e.g. health and safety standards, and unintentional, e.g. global warming). Recent events such as global warming, and the credit crunch highlight the increased complexities in our systems and the interconnections between different geographical locations and entities. For instance, in the case of global warming, as the effect of man-made emissions on climate has been acknowledged, tackling the climate change issue has become a serious priority for governments and international organizations. Although the problem has been recognized, an understanding of how to develop effective, acceptable and detailed policies has yet to be attained. There is a wide debate on the level at which the emission reduction targets should be set, but once values are

agreed, there are a number of possible alternative strategies (policies) to achieve them. Addressing such a complex problem requires the formulation of integrated policies that are well coordinated and reciprocally reinforced. The precise nature and scope of policies designed to achieve their respective targets are necessarily geographically and culturally dependent given the variability of resources, of access to technology and of political constraints at different locations and times. For this reason, a one-for-all and static policy is unlikely to achieve the desired targets and there is a need for bespoke policies that are able to accommodate periodic revisions. The introduction of a systematic approach for exploring alternative policies using a computational methodology will help in decomposing such complex problems into subproblems with more manageable size, accelerate and improve the effectiveness of the process of policy-making. The transport sector has been chosen for the case study because it is the second largest growing source of greenhouse emissions (IPCC, 2007). A new framework for policy formulation is being developed and implemented as a prototype DSS using the backcasting approach around a case study: the formulation and analysis of the policies required to achieve CO_2 emission targets for the UK transport sector.

The background to policy design and backcasting methodology is discussed in Section 2. Section 3 describes the proposed framework for policy design and the details of the DSS. The results achieved in the development of the system are presented in Section 4 followed by a description of future work and the conclusions in Section 5.

2. Background to policy design and backcasting

2.1. Policy Design

A policy is a principle or guideline for action in a specific everyday-world context (Pohl, 2008) and policy design is the step in which the components of a policy are selected and the overall policy is formulated. Policies may be related to technological, economic, political and social aspects. Some technological and economic factors can be modelled mathematically, resulting in complex simulation/prediction models. Mathematical models provide valuable insights, but are only part of the required inputs to a general policy-making process because, after all, decisions about desirable futures, and the policies to attain them, are essentially a question of social values and political choice (Robinson et al., 2006).

The focus within the transport domain has been on the development of mathematical models and tools for assessment of large-scale infrastructure projects and analysis of transport policies. Monetary-based techniques and Multi-Criteria Analysis techniques (especially Multi-Criteria Decision Analysis techniques (MCDA)) form the basis of the models and tools. Often, risk analysis techniques and probabilistic models are used to further refine and fine-tune the models. Both in the case of policy design in general and in the transport sector specifically, the decisions on what to include in the policies (their *synthesis*) is done manually. This fact, together with the size of the space of possible policies, results in a large portion of the design space left unexplored. A systematic approach to explore the space of alternative policies using a computational methodology will accelerate the task of policy-making and improve policy effectiveness.

2.1.1. Process design and policy design

Similar to the transport domain, there are powerful simulators used for chemical process design (Finlayson, 2006). However, in contrast to the transport sector, process engineers have developed a set of tried-and-tested synthesis methods, which help in the selection of the items of equipment and how they are to be interconnected (Westerberg, 2004). In

process design, synthesis and simulation steps are applied in tandem and iteratively: a synthesis step generates alternative designs and the output from a simulation step is used to compare those alternatives and inform the application of the next synthesis step (see Figure 1(a)). The introduction of the aforementioned techniques changed the way in which chemical processes were designed. The plan is to adapt/adopt existing process synthesis and design methods to the design of policies, i.e. to provide practical frameworks and tools for better policy design and support decision-making regarding which measures to use and how to combine them to achieve the targets (the set of tasks included inside the dashed box in figure 1(b)).

(a) Process design
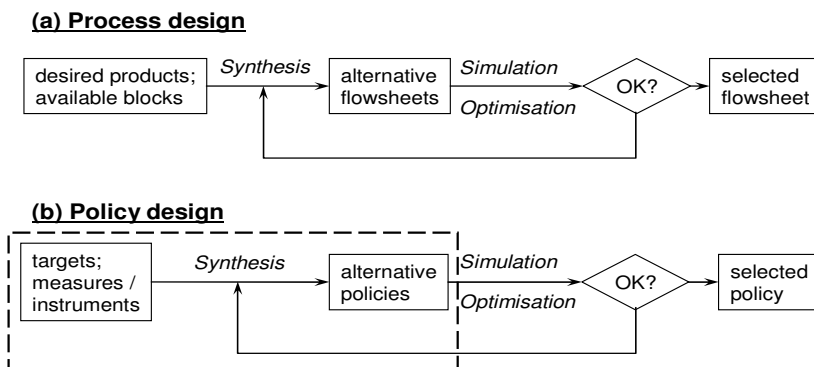
(b) Policy design

Figure 1 Analogy between (a) process design and (b) policy design

The exploration-based model of design (Smithers & Troxell, 1990) and the use of a hierarchical organization are reflected in the widely accepted Hierarchical Design Method (Douglas, 1988), and have been embodied in different design support systems. Our hypothesis is that the process by which the collection of operators is synthesized is similar. However, process and policy design are not identical and, as a result, different approaches will have to be used to take into account their differences, in particular, the pervasiveness of non-quantifiable factors in policy-making.

2.2. Backcasting

Backcasting was first proposed in (Robinson, 1990) and has become a well-established methodology. It involves the development of normative scenarios aimed at achieving desired end-points, i.e. working backwards from a desirable state to determine what policies would be required to reach it. The Visioning and backcasting for UK transport policy project (VIBAT) studied the potential for 60% reduction of emissions due to transport by the year 2030 using a backcasting approach (Banister & Hickman, 2006).

3. A new framework for design of policies

The purpose of the development of the proposed framework and accompanying DSS is to facilitate policy design to achieve environmental targets. The designed system should be applicable to different targets in different sectors and geographical scopes.

3.1. The proposed policy formulation framework is broken down into six steps:

1- Identification of the relevant concepts such as targets and goals.
2- Development of a library of policy measures.
3- Specification of relations among policy measures.

4- Generation of policy packages.

5- Application of planning techniques in the scheduling of policy packages.

6- Evaluation and comparison of alternative policy clusters

3.2. Software architecture and implementation details of the DSS

Figure 2 illustrates the software architecture of the DSS. The connection between the database and the core is achieved through the Java Database Connectivity application programming interface (JDBC, 2008). The Java code currently uses Mathematica's kernel as a computational and visualization engine (Mathematica, 2008). In addition, Mathematica is used for accessing the discrete mathematics package Combinatorica (Pemmaraju & Skiena, 2003). The integration between Java and Mathematica is established using J/Link (J/Link, 2008). The information acquired through user input and analysis via Mathematica is channeled towards the agent-based modeling toolkit. An agent-based model is a system for the simulation of interactions between autonomous entities. Due to the opportunistic nature of agents, it is possible to combine the top-down and bottom-up approaches of design to better address complex problems such as the one under study, allowing for an enhanced utilization of the available information for more effective policy design. Moreover, an agent-based modeling approach is well suited for developing flexible systems that can handle incremental discovery, design and development (North & Macal, 2007) and provides a natural method for decomposing the overall problem.
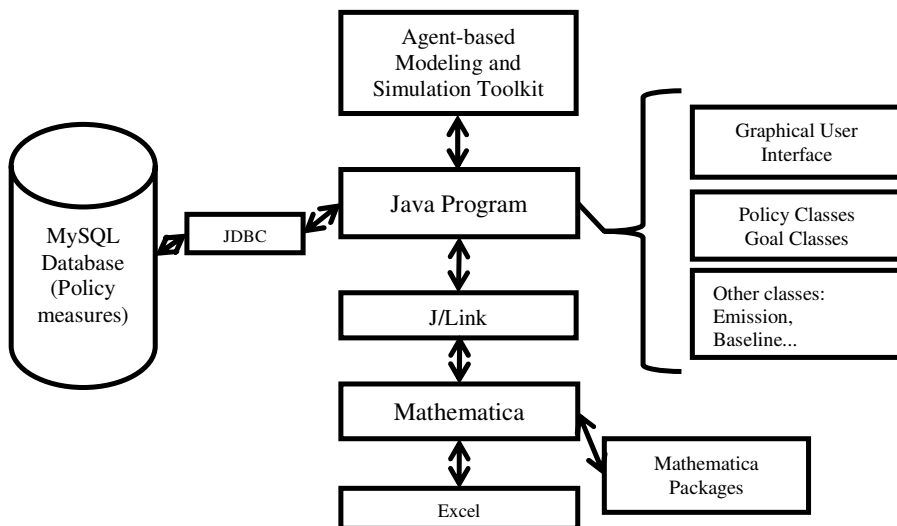
Figure 2 Software architecture of the decision support system

4. Results

4.1. Generation and analysis of the policy measures network

The 122 policy measures from the VIBAT project constitute the core of the library of policy measures with properties such as the policy type, effectiveness and timeframe of implementation. Based on specified preconditions among policy measures, a graph structure was formed with policy measures as nodes (vertices) and the relations as directed edges between the nodes, and several analyses were performed on this network.

Network characteristics, such as degree, betweenness, closeness, and eigenvector centralities were examined. Based on these characteristics, a set of preference indices were created such as *total cost vs. effectiveness* (includes the cost of preconditions), *node cost vs. effectiveness*, *implementation time vs. effectiveness*, and *combined weighted costs and time vs. effectiveness*, which are used later by the agents during the selection and comparison of different policy measures. In addition, graph colouring was used to identify the minimum number of policy packages required for the implementation of all the selected policy measures, also a number of contradictions were defined among policy measures.

4.2. Generation and analysis of the tree of policy alternatives

The next step in the development of the DSS is the generation of desirable combinations of policy measures. A tree structure is utilized for exploring the space of alternatives by generating different policies using the policy measures as building blocks. A multi-layered agent-based architecture is applied, where agents select policy measures as building blocks for creating policy packages, policy clusters and the overall policies. Results from the agent working on creating policy packages are fed to another agent working on policy clusters. This last agent, in turn, creates a policy cluster from the pool of available packages based on its internal rules and interests. In the same manner, results from the policy clusters are fed to the agent which was in charge of creating the overall policies.

The concept of hybrid tree is used to combine goals and policy measures in the same structure. Previously the goals defined by the user were connected to the policy measures through hardwired relations. By formalising the goals, it has been made possible for the agents to connect policy measures to the relevant goals. The goals are formally defined using a method similar to the one proposed by Hashmi et al. (Hashmi, 2003), where each goal has a number of properties such as emission reduction target, a desired time frame for implementation, a level of priority, and a context for the actions. An example goal is to reduce the average level of emissions of the vehicle fleet from 189 g/km in the year 2000 to 90 g/km by the year 2030 with a high priority, which would result in a reduction of 11.8 million tonne of Carbon (MtC) emissions. A set of agents then combine policy measures and goals within a tree structure at different hierarchical levels during the creation of the overall policy.

5. Future Work and Conclusions

The generation and evaluation procedures for the tree structure will be enhanced in the future, and the results from the system will be validated. As it is not possible to compare the results with a real system, since it would require the implementation of the policies beforehand, we intend to compare the results with the VIBAT study and seek transport policy experts' opinion.

It is intended to increase the number of agents that participate in the creation of the tree structure and to use negotiation between agents. Furthermore, general directives such as the DTLR multi-criteria analysis manual (Dodgson, 2001), and tentative guidelines on choosing appropriate multi-criteria decision analysis techniques (Guitouni & Martel, 1998), will be utilized to improve the evaluation of policy packages and clusters.

In the area of network analysis, the effect of the type of relations between nodes with respect to their immediate and non-immediate neighbors will be explored.

The purpose of the research is to facilitate the design of policies by using knowledge gained mainly from process design and synthesis. The focus has been directed towards the similarities between process and policy design with the specific aim of introducing a

new framework and systematic thinking to the problem of policy formulation. A working prototype DSS has been developed; it facilitates the design of transport policies that aim to achieve environmental targets. In particular, the software will help decision makers in selecting appropriate policy measures to achieve a reduction in transport related CO_2 emissions. This research constitutes the first step towards the development of a general family of computer-based systems that support the design of policies to achieve environmental targets for areas such as transport, energy, biofuels/food security, water, etc. The results from the research will be a fresh contribution to the methodological development of policies and have the potential to: (a) accelerate the design of policies. (b) improve the chance of policy success through exploration of more alternatives. (c) facilitate the specialisation of policies.

6. Acknowledgments

References

D. Banister, & R. Hickman, 2006, Visioning and backcasting for UK transport policy (VIBAT), the Bartlett school of planning & Halcrow group ltd, http://www.ucl.ac.uk/~ucft696/vibat2.html

J. Dodgson, M. Spackman, A. Pearman, L. Phillips, 2001, DTLR multi-criteria analysis manual, National Economic Research Associates, UK,   Retrieved 1/18/2008 www.communities.gov.uk/documents/corporate/pdf/146868.pdf

J. M. Douglas, 1988, Conceptual design of chemical processes. McGraw-Hill New York.

B. A. Finlayson, 2006, Process simulation, Chapter 7 in: Introduction to chemical engineering computing. Wiley-Interscience.

A. Guitouni, & J. M. Martel, 1998, Tentative guidelines to help choosing an appropriate MCDA method. European Journal of Operational Research, 109(2), pp. 501-521.

IPCC. 2007, Climate Change 2007: Mitigation. Contribution of the Working group III to the 4th Assessment Report of the Intergovernmental Panel on Climate Change, B. Metz, O.R. Davidson, P.R. Bosch, R. Dave and L.A. Meyer (eds.). Cambridge University Press.

J/Link, 2008, Java toolkit: J/Link: Integrating Mathematica and java. Retrieved 7/8/2008, http://www.wolfram.com/solutions/mathlink/jlink/

JDBC, 2008, The java database connectivity (JDBC) overview, Sun Microsystems inc. Retrieved 7/8/2008 http://java.sun.com/products/jdbc/overview.html

N. Hashmi, A. Boxwala, D. Zaccagnini, J. Fox, 2003, Formal Representation of Medical Goals for Medical Guidelines, Medinfo, pp. 1663.

Mathematica. 2008, Wolfram Mathematica: Home page. Retrieved 7/8/2008 http://www.wolfram.com/products/mathematica/index.html

M.J. North, & C.M. Macal, 2007, Managing business complexity: Discovering strategic solutions with agent-based modeling and simulation. Oxford University Press, New York.

S. V. Pemmaraju, & S. S. Skiena, 2003, Computational discrete mathematics: Combinatorics and graph theory with Mathematica. Cambridge University Press.

C. Pohl, 2008, From science to policy through transdisciplinary research. Environmental Science and Policy, 11(1), 46-53.

J. Robinson, 1990, Futures under glass: A recipe for people who hate to predict. Futures, 22(8), 820.

J. Robinson, M. Bradley, P. Busby, D. Connor, A. Murray, B. Sampson, et al., 2006, Climate change and sustainable development: Realizing the opportunity. Ambio, 35(1), 2-8.

T. Smithers, & W. Troxell, 1990, Design is intelligent behaviour, but what's the formalism? AI EDAM, 4(2), 89-98.

A. W. Westerberg, 2004, A retrospective on design and process synthesis. Computers and Chemical Engineering, 28(4), 447-458.

2055

Continuous-Time Representation Approach to Hybrid Process Scheduling of Single-Product Production

Lijie Su, Lixin Tang

Liaoning Key Laboratory of Manufacturing System and Logistics, The Logistics Institute, Northeastern University, Shenyang, 110004, China

Abstract

This paper investigates the scheduling problem of single-product production on hybrid batch and continuous production processes where the batch process is composed of several parallel reactors and followed by a continuous production line. There is an intermediate storage tank with limited capacity available between two processes. The parallel reactors may simultaneously supply material into the tank. Based on global-event continuous-time modeling method, the scheduling is formulated as a mixed integer nonlinear programming (MINLP) which can be solved by a conventional solver. Two real hybrid processes with certain demand ranges are examined to illustrate the efficiency and applicability of the presented mathematical model.

Keywords: scheduling, hybrid process, global event, continuous-time representation

1. Introduction

The research area of process scheduling has received great attention in the recent 30 years. The complicated process topology structure and variable characters of products make process scheduling different from classical scheduling on discrete manufacture, and mathematical programming approach is often adopted. There are some prominent theoretical modeling fruits in this research field (Méndez, Cerdá, et al 2006). It would be valuable to extend these obtained modeling fruits to more practical production. However, there exists a certain gap between theory methods and practical application for the limitation of methods and the diversity of practical process. For instance, most research work focus on batch process or continuous process, there exist the hybrid process in practice. Hybrid process with batch and continuous modes is one kind of typical processes, which is popular in polymerizing process. Shah, Liberis, et al 1996 presented one integrated approach to design the batch plants. Kang, Kang, et al 2006 formulated the integrated scheduling for the Polyvinyl Chloride (PVC) process. These two references all introduced discrete-time modeling method, and the scheduling was formulated as a mixed integer linear programming (MILP) model where batch processing time is assumed fixed.

Taking PVC process as a research background, we investigate the scheduling of the hybrid process. The new feature of the problem is that scheduling with batching decision is considered and batch processing time is varied. The failure of the discrete-time modeling to tackle the variable batch processing time motivates us to develop the continuous-time method to model the problem under our consideration. Here, global time point is adopted for all units to represent event.

2. Problem Description

The hybrid process here is composed of three parts sequentially: parallel batch reactors, one storage tank and one continuous production line. Multiple tanks and continuous production lines could be regarded as one for the single-product production. And continuous production line could be reduced into one continuous processing unit. Batch size could vary in certain range, and batch processing time is linear function of batch size. The simultaneously unloading operations from multiple reactors to tank are allowed. Once completing batch operation, semi-finished product must transfer into tank at once in order to assure the quality of product. The capacity of tank is limited. The production rate of continuous production line could be adjusted in some range. And the continuous production line would be expected to be running smoothly.

The problem description for the single-product scheduling on one hybrid process is as follow: given the upper and lower bound of the batch size, the variable batch processing time dependent on batch size, the capacity of intermediate storage tank, the production-rate range of continuous production line, total demand quantity of one order. Determine the detail scheduling of the whole process in order to minimize the *makespan*. One assumption is that the delivery durations from reactors to storage tank and storage tank to continuous production line are ignored.

3. Mathematical Formulation for Scheduling Hybrid Process

Concerning with simultaneous multiple-to-one operations from batch units to finite storage tank, global-event continuous-time representation approach is adopted here, which predefines certain number of time points for all units including tank, and any task should happen at some time point (Zhang&Sargent, 1996). Here, the tasks are defined as the finishing operation of batch units and adjusting one of production rate. Compared with unit-specific event based approach (Ierapetritou&Floudas, 1998), global-event one would be more simpler on expression of material balance constraints, and easy to settle storage.

Nomenclature

Indices:

$j = 1, \ldots, J$: batch reactors.

$n = 1, \ldots, N$: event/time points.

Parameters:

D : demand quantity.

$B_j^{\min}, B_j^{\max}$: minimal and maximal capacity of batch reactor j.

α_j, β_j : constant and variable processing time of batch reactor j.

$S^{\max}$: maximal capacity of intermediate storage tank.

$S^{\min}$: minimal required left intermediate product in tank during production process.

$R^{\min}, R^{\max}$: adjusted production-rate range of continuous process.

Variables:

x_{jn} : binary variable, 1 if reactor j completes one batch on time point n, otherwise 0.

y_n : binary variable, 1 if the continuous production process starts working on time point n, otherwise 0.

B_{jn}: continuous variable, batch size which is completed in reactor j on time point n.

P_{jn}: continuous variable, batch processing time of reactor j completed just on time point n.

T_n: continuous variable, the time of the *nth* time point on all units.

SB_n: continuous variable, left intermediate storage quantity at the *nth* time point before receiving new material from reactors.

S_n: continuous variable, intermediate storage quantity at the *nth* time point after receiving intermediate product from some reactors.

R_n: continuous variable, production rate of continuous production process between the *nth* and *n+1th* time point.

MK: continuous variable, the final finishing time of continuous production line, namely, the *makespan*.

The mathematical formulation for the single-product scheduling on the hybrid process is presented as follow.

For the parallel batch reactors, there are allocation, capacity, processing time and time sequence constraints.

There are at most $|J|$ batches completed at any defined time point.

$$\sum_{j=1}^{J} x_{jn} \leq |J| \qquad \forall n = 1...N \tag{1.1}$$

With the balance of work load, the batch number for every reactor should satisfy some limitation.

$$\sum_{n=1}^{N} x_{jn} \leq \left\lceil D \big/ \left(B_j^{\min} \cdot |J| \right) \right\rceil \qquad \forall j = 1...J \tag{1.2}$$

The processed batch amount is bounded by the maximum and minimum capacities of that batch reactor according to reactor capacity and operation safety.

$$x_{jn} \cdot B_j^{\min} \leq B_{jn} \leq x_{jn} \cdot B_j^{\max} \qquad \forall j = 1...J, n = 1...N \tag{1.3}$$

The processing time of one batch is composed by fixed part and variable one, and the later part is dependent on the batch size.

$$P_{jn} = \alpha_j \cdot x_{jn} + \beta_j \cdot B_{jn} \qquad \forall j = 1...J, n = 1...N \tag{1.4}$$

The time of the *nth* time point is less than that of the *n+1th* time point.

$$T_n \leq T_{n+1} \qquad \forall n = 1...N - 1 \tag{1.5}$$

For any reactor j, the time of time point n must be larger than the summation of all batches processing time just before time point n.

$$\sum_{m=1}^{m=n} P_{jm} \leq T_n \qquad \forall j = 1...J, n = 1...N \tag{1.6}$$

For the storage tank, there are material balance and capacity constraints.

The left intermediate-product quantity on time point n equals to the intermediate storage quantity on *n-1* minus the quantity processing by continuous production line.

$$SB_n = S_{n-1} - R_{n-1} \cdot \left(T_n - T_{n-1} \right) \qquad \forall n = 2...N \tag{1.7}$$

The intermediate storage quantity on time point n is the left intermediate-product quantity on time point n plus completed batch quantity from reactor at the same time point.

$$S_n = SB_n + \sum_{j=1}^{J} B_{jn} \qquad \forall n = 2...N \qquad (1.8)$$

The reason for using two variables defining the storage quantity on time point n is that consumption of continuous production has been lasted between the *n-1th* and *nth* time points, however, production of batch reactor happens at the *nth* time point. Two material balance equations would constraint the consumed quantity was not the just produced one at time point n. For the first time point, we use the following constraint

$$S_1 = SB_1 = \sum_{j=1}^{J} B_{j1} \qquad (1.9)$$

It is the limit on maximal intermediate storage quantity on any time point.

$$S_n \leq S^{\max} \qquad \forall n = 1...N \qquad (1.10)$$

In order to ensure the continuity of next procedure, some minimal quantity of intermediate product must be remained in storage tank.

$$SB_n \geq S^{\min} \qquad \forall n = 1...N \qquad (1.11)$$

This constraint would make the gotten schedule more flexible, for the minimal quantity of intermediate product could be regarded as a buffer between batch reactors and continuous production line.

For the continuous process part, there are processing rate, smoothing production and *makespan* constraint.

This is the adjusted range of production rate when continuous process runs smoothly.

$$y_n \cdot R^{\min} \leq R_n \leq y_n \cdot R^{\max} \qquad \forall n = 1...N \qquad (1.12)$$

The followed constraint would assure the uninterrupted production for continuous production line, which means once the continuous production process starts up, it would continue until final completion.

$$y_n \geq y_{n-1} \qquad \forall n = 2...N \qquad (1.13)$$

The completion time of one order with certain demand, the *makespan,* is greater than the summation of the time of last time point N plus the processing time of quantity S_N.

$$MK \geq T_N + S_N / R^{\max} \qquad (1.14)$$

The total production quantity must satisfy demand. Here, D means the quantity of intermediate product converted from one of final product.

$$\sum_{j=1}^{J} \sum_{n=1}^{N} B_{jn} \geq D \qquad (1.15)$$

The objective function is minimization of makespan, that is

$$\min \quad MK \qquad (1.16)$$

Proposition 1. For one certain demand quantity, the upper bound of batch number is $\lceil D/B^{\min} \rceil$, and the lower bound of it is $\lceil D/B^{\max} \rceil$.

Here, $B^{\min} = \min_{j} \left\{ B_j^{\min} \right\}$, $B^{\max} = \max_{j} \left\{ B_j^{\max} \right\}$.

According to proposition 1, the valid cut inequalities would be presented.

$$\sum_{j=1}^{J}\sum_{n=1}^{N} x_{jn} \le \left\lceil \frac{D}{B^{\min}} \right\rceil \tag{1.17}$$

$$\sum_{j=1}^{J}\sum_{n=1}^{N} x_{jn} \ge \left\lceil \frac{D}{B^{\max}} \right\rceil \tag{1.18}$$

Proposition 2. The lower bound for the problem *makespan* is

$$MK^{LB} = \min_{j}\left\{\alpha_j + \beta_j \cdot B_j^{\min}\right\} + D/R^{\max}. \tag{1.19}$$

The shortest duration for completing demand is $D/R^{\max}$, if continuous production line starts at the beginning of horizon. Concerning batch processing time, the earliest time getting intermediate product is $\min_{j}\left\{\alpha_j + \beta_j \cdot B_j^{\min}\right\}$, as the condition for the beginning of continuous production line. Therefore, the idea objective for the whole production process is the summation of the two parts.

The lower bound valid inequality is

$$MK \ge MK^{LB} \tag{1.20}$$

4. Case Studies

All the following data from process structure to processing time is gotten from real PVC production processes. The number of parallel units is 4 and 8, which operational capacity varies from 35m³ to 39m³. The batch processing time is calculated by $\alpha_j = 22, \beta_j = 0.05 \quad \forall j = 1, \cdots, J$. Other data would be within each case.

LINGO 8.0 is used to solve the MINLP formulation, and the selected solver is Branch and Bound. In order to accelerate convergence, higher level of heuristics and probing strategies are set. The used computer is Intel Core 2 Due CPU, and the CPU is 2.33 GHz. And the operating system is Windows XP professional 2002.

Proposition 3. The least number of predefined time points is $\left\lceil D/\left(S^{\max} - S^{\min}\right) \right\rceil$.

From the least number of predefined time points, iterative procedures would be applied until the value of the optimal solution not decreases, or decreases little.

The hybrid process I includes 4 reactors in batch part. The capacity of storage tank is 80 ton, the minimal storage quantity of intermediate product is 5 ton, and the production rate varies from 3.6 t/h to7.2 t/h. The computational results are listed in Table 1.

Table 1 Statistic results for Hybrid process I

Demand	Events	Int vars	Cont vars	Lower bound	Solution	CPU s
300	5	25	143	65.4	68.4	3
350	6	30	170	72.4	81.7	38
400	7	35	197	79.3	86.5	19
500	9	45	251	93.2	105.4	113
600	10	50	278	107.1	115.9	103
800	11	55	305	134.9	152.9	112
900	14	70	386	148.8	163.3	104
1000	16	80	440	162.6	181.5	117
1500	23	115	629	232.1	258.9	110
2000	28	140	764	301.5	343.9	129

The hybrid process II includes 8 reactors in batch part. The capacity of storage tank is 160 ton, the minimal storage quantity of intermediate product is 4 ton, and the production rate varies from 7.2 t/h to 14.4 t/h. The demand changes from 500 ton to 3000 ton. The computational results are listed in Table 2.

Table 2 Statistic results for Hybrid process II

Demand	Events	Int vars	Cont vars	Lower bound	Solution	CPU s
500	5	45	227	58.5	61.0	55
800	7	63	313	79.3	84.2	180
1000	7	63	313	93.2	103.1	114
1200	8	72	356	107.1	114.2	131
1500	11	99	485	127.9	137.8	114
1800	16	144	700	148.8	164.4	115
2000	16	144	700	162.6	177.5	109
2200	16	144	700	176.5	196.2	105
2500	20	180	872	197.4	221.8	177
3000	24	216	1044	232.1	258.8	170

Here, the solutions in table 1-2 are local optimums gotten by LINGO.

Observing the statistic results of numerical experiments, we find that the proper number of time points usually is near to the least number of predefined time points. The objective function reduces little with the increase of time points, the changes are negligible and multiple local optimal solutions are supplied, which would be flexible for the production management. Comparing with lower bounds, the local optimal solutions are good enough. And the CPU running time is sustainable.

5. Conclusion

This work introduced the global-event based formulation to the hybrid process scheduling. The formulation is applied to the hybrid process of flexible batch size and batch processing time, simultaneous multiple-to-one operations, finite storage, adjusted production rate. The numerical experiments validate the efficiency and practicability of the formulation, which could be extended into the similar practical production process with little modifications. The presented model could also be used to optimize the capacity of storage and maximal constant production rate for certain demand within given time horizon.

References

M. G. Ierapetritou and C. A. Floudas, 1998, Effective Continuous-Time Formulation for Short-Term Scheduling. 1. Multipurpose Batch Processes, Industrial and Engineering Chemistry Research, Vol. 37, 4341-4359.

M. G. Kang, S. Kang, and S. W. Park, 2006, Integrated Scheduling for Polyvinyl Chloride Processes, Industrial and Engineering Chemistry Research, Vol. 45, 5729-5737.

C. A. Méndez, J. Cerdá, I. E. Grossmann, I. Harjunkoski, M. Fahl, 2006, State-of-the-art Review of Optimization Methods for Short-term Scheduling of Batch Processes, Computers and Chemical Engineering, Vol. 30, 913–946.

N. Shah, L. Liberis, E. Izumoto, R. Henson, 1996, Integrated Batch Plant Design: A Polymer Plant Case Study, Computers and Chemical Engineering. Vol. 20, S1233-S1238.

X. Zhang, R. W. H. Sargent, 1996, The Optimal Operation of Mixed Production Facilities-A General Formulation and Some Solution Approaches for the Solution, Computers and Chemical Engineering, Vol. 20, 897-904.

Evaluation of Synergy Effect in the Merger of Companies in a Petrochemical Complex

Sung-Geun Yoon[a], Sunwon Park[a], Jeongseok Lee[b], Peter M. Verderame[c], and Christodoulos A. Floudas[c]

[a]*Department of Chemical and Biomolecular Engineering, KAIST, 373-1, Guseong-dong, Yuseong-gu, Daejeon, 305701, Korea*
[b]*Corporate R&D, LG Chem Ltd., 104-1, Moonji-dong, Yuseong-gu, Daejeon, 305380, Korea*
[c]*Department of Chemical Engineering, Princeton University, Princeton, New Jersey, 08544, U.S.A.*

Abstract

Mergers and Acquisitions (M&A) have been actively carried out in the petrochemical industry. However, the synergy created by the merger of petrochemical companies has seen relatively little study, despite being the primary goal of a merger. This study deals with the horizontal merger of petrochemical companies located within a single complex. Synergies considered in this paper stem from integration of the process network and the utility plant, fixed cost reduction, and contracts in purchasing and selling. A novel mathematical model that represents the operation of a process network and a utility plant and the decision for purchasing and selling contracts is formulated. Four contracts for purchasing and selling are considered. The proposed model is applied to three Korean Naphtha Cracking Center (NCC) companies located in the same industrial complex. The results show that synergy effects from integration of the process network and the utility system, fixed cost reduction, and increased market share together increase profit by fifty percent.

Keywords: Mergers and acquisitions, synergy, petrochemical company

1. Introduction

Recently, the Korean petrochemical industry has experienced prosperity due to rapid growth of the neighbouring Chinese market. That China and Middle Eastern countries are rapidly developing their own petrochemical industries, however, poses a major threat to the Korean petrochemical industry. In order to sustain competitiveness, the Korean petrochemical industry has recently come to consider M&A as an important strategy. At present, too many companies are participating in the petrochemical industry. Companies and government should therefore consider M&A as a necessary strategy to improve competitiveness.

However, there is no blueprint for M&A strategies from a holistic view. Some companies have announced mergers of their subsidiary companies. The government also has not provided direction regarding M&A. Many stakeholders are now debating on the M&A issue. The debate has focused on how much M&As will impact the petrochemical industry and the national economy. In order to assist stake holders, in an earlier study the present authors proposed an optimization model to quantify synergy in

the merger of petrochemical companies in an industrial complex, considering purchasing and selling advantages (Yoon et al., 2008). This paper deals with the same target companies that were considered in our previous work. We propose a new and improved optimization model to reflect more realities. The model can quantify synergy from the case in which a target company has one utility plant respectively. Investment cost to enable transport of chemicals and steam between individual companies is added to the optimization model. The single period model is extended to a multi-period model. And four representative contracts for purchasing and selling are applied to the model.

2. Problem Definition

This paper deals with three companies, A, B, and C, having a naphtha cracking center (NCC) within a complex in Korea. Each company has NCC and BTX processes and different downstream processes. Each company has one utility plant having one boiler, one SS steam turbine, and one HP steam turbine. The capacities of the boilers and turbines are different. There are steam and electricity transportations between the process network and the utility plant. Some processes generate steam, which is transported to the utility plant. The utility plant produces steam and electricity and then supports them to the process network according to the demands of each steam grade and electricity. NCC companies can sell steam to other companies. Sometimes NCC companies sell electricity generated by their steam turbine to a regional power company.

If the three companies are merged, the new merged company can operate all process networks and utility plants of the three companies. We assume that company D is the aggregate of companies A, B, and C. The process network and utility plant diagram of company D are shown in Figure 1.
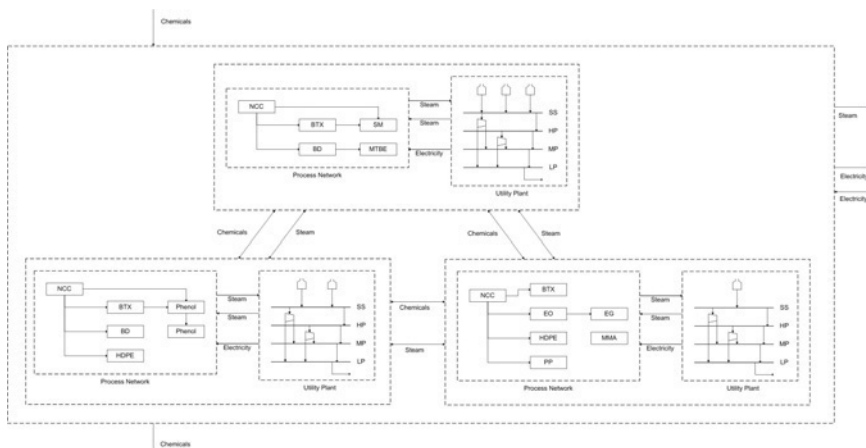
Figure 1. Process network and utility plant diagram of Company D

3. Mathematical Modeling

A novel MILP model is developed for optimization of a process network and a utility plant. The model maximizes the profits of target companies under given price and parameter data. The nomenclature is presented in Appendix.

3.1. Objective Function

$$\text{Maximize } z = \sum_t \sum_h \sum_j REV_{hjt} - \sum_t \sum_h \sum_j PUR_{hjt} - \sum_t \sum_n \sum_k FP_{nkt} \cdot fuel_{nkt}$$
$$- \sum_t \sum_n \sum_i (U_{nit}(0.4 \cdot b_{nit} + 0.6 \cdot r_{nit}) + F_{nit}) \overline{X_{nit}} - \sum_t PE_t \cdot ep_t + \sum_t SE_t \cdot es_t$$
$$+ \sum_t \sum_n \sum_o stms_{not} \cdot stmp_{ot} - \sum_t \sum_n \sum_{n'} \sum_j invct_{n\,n'\,jt} - \sum_t \sum_n \sum_{n'} \sum_o invst_{n\,n'\,ot}$$

(1)

The objective is to maximize the net present value of the profit. A company intends to operate their plant optimally based on a given business condition. The profit is calculated by subtracting chemical purchasing, operating cost and investment cost from the chemical sales.

3.2. Purchasing and Selling

$$Qp_{jt} = \sum_n nqp_{njt}, \ \forall j,t \ (2), \quad Qs_{jt} = \sum_n nqs_{njt}, \ \forall j,t \ (3)$$

Equations (2) and (3) calculate amount of chemical purchasing and selling, Qp_{jt} and Qs_{jt}, of the merged company. The types of contracts for purchasing include (1) fixed price, (2) discount after a certain amount, (3) bulk discount, and (4) fixed duration. Equations to determine purchasing cost for each type of contract are given by:

$$PUR_{jt} = \sum_h PUR_{hjt}, \ \forall j,t \ (4), \quad Qp_{jt} = \sum_h qpc_{hjt}, \ \forall j,t \ (5)$$

$$0 \leq qpc_{hjt} \leq M \cdot bqpc_{hjt}, \ \forall h,j,t, \ M = \text{large positive} \ (6), \quad \sum_h bqpc_{hjt} \leq 1, \ \forall j,t \ (7)$$

Equation (4) means that PUR_{jt}, purchasing cost of chemical j at time t, is defined as the sum of PUR_{hjt}. Equation (5) represents that Qp_{jt}, purchasing amount of chemical j at time t, is the sum of qpc_{hjt}, purchasing amount of chemical j in contract h at time t. Equation (6) activates binary variable for qpc_{hjt}. Large M used in this study is 10^6. The number of contracts that can be made at time t is constrained by equation (7). In the same way to the purchasing contracts, a company can sell chemicals through different types of contracts. The types of contracts for selling include (1) fixed price, (2) discount after a certain amount, (3) bulk discount, and (4) fixed duration. Detailed models for contracts, operation of process network and utility system, and investment are refereed to works of Park et al. (2006) and Yoon et al. (2008, 2009).

4. Optimization Study

In this section, we apply the proposed mathematical model to individual companies, A, B and C and a merged company D. Synergy of merger is defined as difference between profit of company D and the sum of profits of companies A, B and C. The case study is based on five years time periods. An interest rate used in the optimization is five percent. Table 1 shows model statistics for the optimization study.

Optimization results show the NPV of profits of three individual companies and a merged company. Synergy is calculated using the results. Table 2 shows the results and synergy. The results exhibit that the merger of three NCC companies creates synergy of 1,851 million dollars. The synergy increases the profit by almost fifty percent.

Investment for chemical and steam transportation is 2.93 million dollars that is much smaller than the synergy created.

Table 1. Model statistics for the optimization study

Company	Equations	Continuous variables	Binary variables	Solution time (s)	Iterations
A	24158	18936	6905	0.406	618
B	24158	18936	6905	0.343	368
C	24158	18936	6905	0.234	397
D	49773	36826	11860	3.000	5220

Table 2. Optimization results

Profit (mil. $)	Individual Companies	A	1,333.93
		B	1,093.84
		C	1,305.13
		Sum	3,732.89
	Merged Company	D	5,584.36
Synergy (mil. $)			1,851.46
Synergy Effect (%)			49.60
Investment (mil. $)			2.93
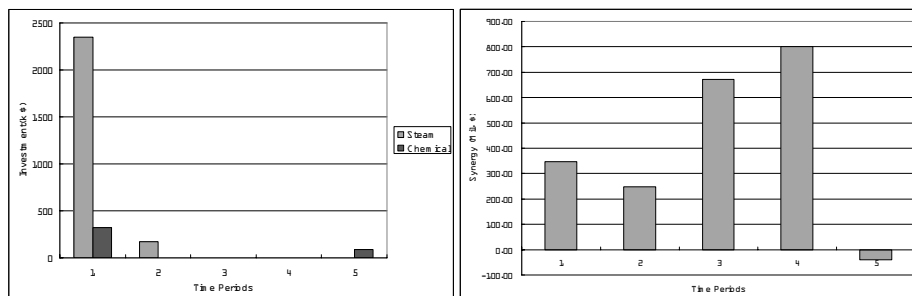
Figure 2. Investment cost (left) and Synergies in each time period (right)

A merger needs rationalization of production facility. For petrochemical companies, the rationalization means integration of process networks and utility systems. The integrations require capital investment for chemical and steam transportations. Figure 2 (left) shows investment cost for chemical and steam transportations. Investment for steam transportation is much larger than investment for chemical transportation. Most investment is made in time period 1.

Synergies in each time period are shown in Figure 2 (right). Except in time period 5, positive synergy is created in every time period. This study considers four kinds of synergy factors: process network integration, utility system integration, fixed cost reduction, and contracts for purchasing and selling. Contributions of four factors to the synergy need to be analyzed. In order to identify contributions of the factors, optimization studies are carried out without fixed cost reduction and chemical and steam transportations respectively. The chemical transportation represents process network integration to create synergy. The steam transportation represents utility system integration to create synergy. The contribution of contracts is calculated by subtracting

the contributions of fixed cost reduction, process network integration and utility system integration from the total synergy. Table 3 shows results of the optimization studies and contribution of each factor.

Table 3. Contributions of four factors (Mil. $)

Profit of the merged company, D	5584.36
without fixed cost reduction	5427.23
without process network integration	5423.44
without utility system integration	5538.68
Contributions of	
Fixed cost reduction	157.13 (8.49 %)
Process network integration	160.92 (8.69 %)
Utility system integration	45.68 (2.47 %)
Contracts	1487.74 (80.35 %)

The results represent that contracts in purchasing and selling contribute the most to synergy. Contract affects the revenue and the purchasing cost. Figure 3 shows that the amount of naphtha purchasing in Company D is larger than the sum of the amounts of individual companies. Unit naphtha purchasing cost of D is lower than those of individual companies on average. Selling amount of ethylene of D is larger than the sum of those of A, B and C in Figures 4. Unit selling price of ethylene, however, is almost the same in A, B, C and D. In results, the larger purchasing amount of naphtha in the merged company leads to a better price of the naphtha and then, produces more products.

5. Conclusion

This paper proposes the novel mathematical model to quantify various synergies in the merger of petrochemical companies within a complex. We focus on synergy from contracts in purchasing and selling, process network integration, utility system integration and fixed cost reduction. The investment term is included to calculate the cost to make chemical and steam transportation real. The proposed model integrates planning of the contracts for purchasing and selling and operation of the process network and utility system with M&A of petrochemical companies. Specific planning for post merger situation leads to precise estimation of the synergy. The optimization study is carried out on Korean NCC companies. The results show that the merger increases profit by fifty percents. The purchasing contract contributes the most to creating synergy. The larger purchasing amount leads to lower unit cost of naphtha. Korean petrochemical industry seriously needs to consider the merger as a growth strategy.

References

S. Yoon, SB. Park, S. Park, J. Lee, PM. Vederame, CA. Floudas, 2008, Synergy in Mergers of Petrochemical Companies within a Complex Considering Purchasing and Selling Advantage, Ind. Eng. Chem. Res., 47, 5556

M. Park, FD. Mele, S. Park, 2006, Modeling of Purchase and Sales Contracts in Supply Chain Optimization, Ind. Eng. Chem. Res., 45, 5013

S. Yoon, SB. Park, S. Park, J. Lee, PM. Vederame, CA. Floudas, 2009, Selecting the Optimal
    Target Company Based on Synergy Calculation for the Vertical Merger in a Petrochemical
    Complex., Ind. Eng. Chem. Res., 48, 1511
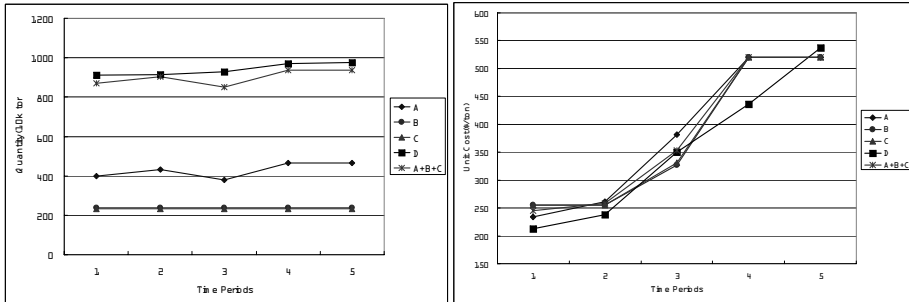
Figure 3. Naphtha purchasing amount (left) and unit purchasing cost (right)
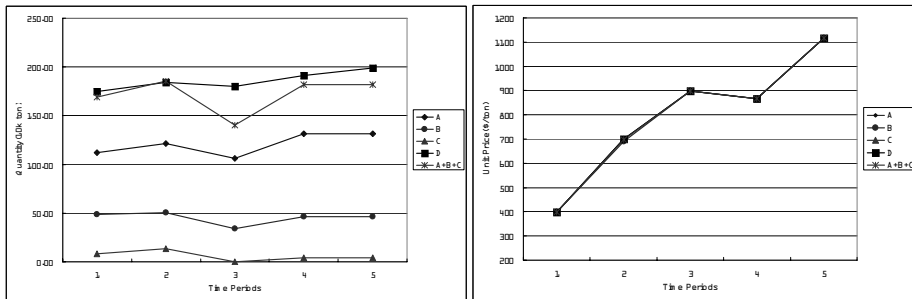
Figure 4. Ethylene selling amount (left) and unit selling price (right)

Appendix: Nomenclature

Sets

n: Company, i: Process, j: Chemical, k: Boiler, h: Contract, o: Steam grade, t: Time

Parameters

ep_t, es_t: electricity purchasing and selling price ($/MWh)

F_{nit}, U_{nit}: fixed cost and and utility cost for process ($/ton)

FP_{nkot}: fuel price of boiler ($/toe)

$stmp_{ot}$, $stms_{not}$: steam selling price and amount ($/ton)

$\overline{X}_{nit}$: process capacity (ton/yr)

Variables

b_{nit}: binary variable for process operation

$fuel_{nkot}$: fuel consumption in boiler (toe/yr)

$invct_{nn'jt}$, $invst_{nn'ot}$: investment cost for chemical and steam transportations ($)

PE_t, SE_t: electricity purchasing and selling (MWh/yr)

PUR_{hjt}, REV_{hjt}: purchasing cost and sales revenue of chemical in contract h ($/yr)

r_{nit}: operating ratio of process (ton/yr)

2067

Optimal Assignment of Plant Operators on Basis of Shift's Ability Evaluation

Masaru Noda and Hirokazu Nishitani

Graduate Scool of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0192, Japan

Abstract

A mathematical programming approach is proposed for solving a shift operator assignment problem on the basis of an evaluation of each shift's abilities. To evaluate the abilities of the workers of a shift, the minimum required skill levels for all processes are defined for three plant conditions: normal situations, abnormal situations, and emergencies. The ratio of successful combinations to all the assumed combinations of abnormal processes is defined as an index of the shift's ability to cope with the situation. The optimal assignment of operators to each shift is formulated as a nonlinear integer programming problem, where the minimum ability among all shifts is maximized. The optimal solution of the operator assignment problem shows us the limitation of operations by the assigned shift members under the specified situation. The proposed approach is helpful for managing abnormal situations and for improving operator training.

Keywords: Operator assignment, Multi-skilled operator, Shift ability evaluation, Abnormal situation management

1. Introduction

Most plants in the chemical industry are operated continuously by shifts that are formed with a certain number of operators. Each shift is composed of multi-skilled operators who have different skill levels for different processes. As a result, when the condition of a process becomes abnormal, a higher-skilled operator on the shift who can cope with the abnormality must temporarily take the place of a lower-skilled operator. Thus, for abnormal situation management, it is necessary to quantitatively evaluate a shift's ability to cope with plant abnormalities.

Previously, we proposed a quantitative method to evaluate the abilities of a shift under specified abnormal plant conditions (Noda *et al.*, 2008). We defined an index for a shift's ability to cope with abnormal situations by the ratio of successful combinations to all possible combinations of abnormal processes. We can find bottlenecks of the shift by investigating all of the patterns that the shift cannot cope with. In order to resolve these bottlenecks, we need to optimally assign operators to shifts on the basis of the shift's ability evaluation or to improve the skill levels of operators.

A human resource management problem in an organization operated by shifts had been discussed by many researchers. For example, Ikegami *et al.* (2003) proposed a new mathematical model and an approach to the nurse scheduling problem. In this study, we formulate a new mathematical programming problem for solving a shift operator assignment problem in a plant on the basis of the shift's ability evaluation. In the formulation, the minimum ability among all shifts is maximized by optimally assigning

operators to shifts. The proposed method is useful for abnormal situation management that ensures safe plant operations under specified abnormal situations.

2. Evaluation method of shift's ability

2.1. Definition of operator's skill level

A shift is composed of multi-skilled operators, who have different skill levels for different processes. In order to assess the ability of a shift to cope with abnormal plant conditions, we first need to quantitatively evaluate the skill level of each operator in each process. The definition of skill levels of course depends on the company and the plant (Takeda, 2002), but in this study, we use the following five skill levels.

> Skill level 1: Beginner
> Skill level 2: Able to do routine operations
> Skill level 3: Able to cope with minor problems
> Skill level 4: Able to cope with abnormal conditions
> Skill level 5: Able to cope with emergencies

2.2. Formulation of minimum requirements

To quantitatively assess the ability of a shift to cope with a specified abnormal situation, we formulate a linear integer programming problem, where a typical continuous chemical plant with several processes operated by different shifts is used as an example to explain the formulation. In this plant, there are N_p processes, and each process is operated by one operator. Thus, the total number of operators in a shift is N_p. We define the minimum required skill level for all processes under three types of plant conditions: normal, abnormal, and emergency. The minimum skill levels required for operating normal, abnormal, and emergency processes are 2, 4, and 5, respectively.

To formulate an optimal shift assignment problem, we introduce an integer variable $z(i, p)$. When operator i is assigned to process p, $z(i, p) = 1$. When operator i is not assigned to process p, $z(i, p) = 0$. Because each operator can be assigned to only one process, the following constraint must be satisfied for all operators:

$$\sum_{p=1}^{N_p} z(i, p) = 1 \quad \text{for all } i, \tag{1}$$

and because each process is operated by one operator, the following constraint must be satisfied for all processes:

$$\sum_{i=1}^{N_p} z(i, p) = 1 \quad \text{for all } p. \tag{2}$$

To quantitatively assess the ability of a shift to cope with a specified situation, we introduce the objective function given by Eq.(3), where $c(i, p)$ is a cost coefficient calculated by Eq.(4). Here, $s(i, p)$ is the skill level of operator i for process p, and $e(p)$ is the minimum skill level required for operating process p. Eq. (3) denotes the total sum of deficiencies of operator's skill levels from satisfying the minimum requirements for all processes.

$$\sum_{p=1}^{N_p} \sum_{i=1}^{N_p} c(i, p) \cdot z(i, p) \tag{3}$$

$$\left. \begin{array}{ll} c(i, p) = 0, & \text{if } s(i, p) \geq e(p) \\ c(i, p) = e(p) - s(i, p), & \text{if } s(i, p) < e(p) \end{array} \right\} \tag{4}$$

When the skill levels of all operators in a shift and all combinations of abnormal processes are given, an optimal assignment problem of shift operators can be formulated as the following linear integer programming problem (LIP):

(LIP)
Minimize Eq. (3)
subject to Eqs. (1) and (2).

If the value of the objective function for the optimal solution is zero, the shift can satisfy the minimum requirements. When it is not zero, the shift is unable to cope with the specified situation.

2.3. Shift's ability evaluation

The LIP can be used to find combinations of abnormal processes that the shift cannot cope with. The ratio of combinations of abnormal processes with which a shift can cope to all the assumed combinations of abnormal processes is used as an index of the shift's ability to cope with abnormal situations. When this index is 100%, the shift can cope with all the assumed combinations of abnormal processes.

In this study, the following three evaluation indices of a shift's ability are defined by Eqs. (5) – (7).

$$A_n = \frac{N'_n}{N_n} \times 100 \ [\%] \tag{5}$$

$$A_a = \frac{N'_a}{N_a} \times 100 \ [\%] \tag{6}$$

$$A_e = \frac{N'_e}{N_e} \times 100 \ [\%] \tag{7}$$

Where, A_n, A_a, and A_e are evaluation indices of the shift's abilities under normal, abnormal, and emergency plant situations, respectively, and N'_n, N'_a, and N'_e indicate the total numbers of combinations of normal, abnormal, and emergency processes with which a shift can cope. The values N_n, N_a, and N_e represent the total numbers of all the assumed combinations of normal, abnormal, and emergency processes. The possible combinations of abnormal and emergency processes should be determined on the basis of the history of problems and on the characteristics of each process. If the possible combinations of processes are unknown, it may be worth investigating the maximum number of processes in the plant that the shift can handle safely.

3. Optimal assignment of operators based on shift's ability evaluation

We consider an integrated control room, where the plant is operated by N_g shifts. There are N_p processes in a plant, and each process is operated by one operator. Thus, the total number of operators in a shift is N_p. To formulate a mathematical programming problem for solving a shift operator assignment problem on the basis of the shift's ability evaluation, we introduce the integer variable $x(i, g)$ as an optimization variable. When operator i is assigned to shift g, $x(i, g) = 1$. When operator i is not assigned to shift g, $x(i, g) = 0$. Because each operator can be assigned to only one shift, the following constraint must be satisfied for all operators:

$$\sum_{g=1}^{N_g} x(i,g) = 1 \quad \text{for all } i . \tag{8}$$

Furthermore, because the total number of operators in each shift is N_p, the following constraint must be satisfied for all shifts:

$$\sum_{i=1}^{N_i} x(i,g) = N_p \quad \text{for all } g , \tag{9}$$

where N_i is the total number of operators in the plant. Because all the shifts have to cope with a normal operating condition, the following constraint must be satisfied for all shifts.

$$A_n(g) = 100 \quad \text{for all } g \tag{10}$$

When the minimum shift's ability in abnormal situations is given from the viewpoint of safe plant operations, Eq. (11) is introduced in the optimization problem as a constraint.

$$A_a(g) \ge A_{a,\min} \quad \text{for all } g \tag{11}$$

Here, $A_{a,\min}$ denotes the minimum limit for A_a.

In this study, the minimum ability in emergencies among all shifts is maximized by optimally assigning operators to all shifts. The operator assignment problem is newly formulated as a nonlinear integer programming problem (NIP).

> (NIP)
> Maximize (min $A_e(g)$)
> subject to Eqs. (8), (9), (10), and (11).

4. Case study

4.1. Shift's ability evaluation

The proposed method is demonstrated through a case study, where a chemical plant with seven processes is operated by shifts, 1 and 2. Each shift consists of seven operators, and each process is operated by one operator. Thus, the total number of operators in this plant is fourteen. The skill level of operator i for process p is listed in Table 1. Fourteen operators were assigned to two shifts so that average skill levels of operators in the shifts became nearly equal (3.4).

Table 1 Skill levels of all operators in shifts 1 and 2

$i \backslash p$	Shift 1							$i \backslash p$	Shift 2						
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
1	5	5	5	5	5	4	4	8	4	4	4	4	5	5	5
2	5	5	4	4	3	4	4	9	5	5	4	4	3	4	4
3	4	4	3	3	1	4	4	10	4	4	3	3	3	5	5
4	4	4	3	4	1	3	3	11	4	4	3	4	3	4	4
5	4	4	3	4	1	3	3	12	4	4	3	4	1	1	1
6	4	4	3	4	1	1	1	13	4	4	4	4	1	1	1
7	4	4	3	4	1	1	1	14	3	3	3	3	1	1	1

First, the abilities of these two shifts were evaluated for the three plant conditions: normal, abnormal, and emergency situations. In the normal situation, all processes in

the plant are normal. In an abnormal situation, it is assumed that three processes become abnormal simultaneously. In an emergency, it is assumed that emergency conditions occur in two processes simultaneously. The total numbers of patterns for normal, abnormal, and emergency situations are 1, 35 ($_7C_3$), and 21 ($_7C_2$) respectively, and are listed in Tables 2, 3, and 4. In the tables, 0 means that the process is normal, 1 means that the process is abnormal, and 2 means the process is in an emergency. The minimum skill levels required for operating normal, abnormal, and emergency processes are 2, 4, and 5 respectively.

Table 2 Pattern of normal plant condition

No	P1	P2	P3	P4	P5	P6	P7
1	0	0	0	0	0	0	0

Table 3 Patterns of abnormal conditions

No	P1	P2	P3	P4	P5	P6	P7
1	0	0	0	0	1	1	1
2	0	0	0	1	0	1	1
3	0	0	0	1	1	0	1
4	0	0	0	1	1	1	0
5	0	0	1	0	0	1	1
6	0	0	1	0	1	0	1
7	0	0	1	0	1	1	0
8	0	0	1	1	0	0	1
9	0	0	1	1	0	1	0
10	0	0	1	1	1	0	0
11	0	1	0	0	0	1	1
12	0	1	0	0	1	0	1
13	0	1	0	0	1	1	0
14	0	1	0	1	0	0	1
15	0	1	0	1	0	1	0
16	0	1	0	1	1	0	0
17	0	1	1	0	0	0	1
18	0	1	1	0	0	1	0
19	0	1	1	0	1	0	0
20	0	1	1	1	0	0	0
21	1	0	0	0	0	1	1
22	1	0	0	0	1	0	1
23	1	0	0	0	1	1	0
24	1	0	0	1	0	0	1
25	1	0	0	1	0	1	0
26	1	0	0	1	1	0	0
27	1	0	1	0	0	0	1
28	1	0	1	0	0	1	0
29	1	0	1	0	1	0	0
30	1	0	1	1	0	0	0
31	1	1	0	0	0	0	1
32	1	1	0	0	0	1	0
33	1	1	0	0	1	0	0
34	1	1	0	1	0	0	0
35	1	1	1	0	0	0	0

Table 4 Patterns in emergencies

No	P1	P2	P3	P4	P5	P6	P7
1	0	0	0	0	0	2	2
2	0	0	0	0	2	0	2
3	0	0	0	0	2	2	0
4	0	0	0	2	0	0	2
5	0	0	0	2	0	2	0
6	0	0	0	2	2	0	0
7	0	0	2	0	0	0	2
8	0	0	2	0	0	2	0
9	0	0	2	0	2	0	0
10	0	0	2	2	0	0	0
11	0	2	0	0	0	0	2
12	0	2	0	0	0	2	0
13	0	2	0	0	2	0	0
14	0	2	0	2	0	0	0
15	0	2	2	0	0	0	0
16	2	0	0	0	0	0	2
17	2	0	0	0	0	2	0
18	2	0	0	0	2	0	0
19	2	0	0	2	0	0	0
20	2	0	2	0	0	0	0
21	2	2	0	0	0	0	0

The evaluation results of the shift's ability under the three plant conditions for two shifts in Table 1 are summarized in Table 5. The evaluation results indicate that the abilities of the two shifts are high in normal and abnormal conditions, but very low in

emergencies. The ability of shift 1 in an emergency is lower than that of shift 2, although the average skill levels of the two shifts are equal. The reason for the low ability of shift 1 in an emergency is due to an unbalanced assignment of operators to the shifts. In shift 1, only two operators have skill level 5 and no operator can cope with an emergency in process 6 or 7.

Table 5 Shift abilities in three plant conditions

Shift no.	Operator's no.	A_n	A_a	A_e
1	1,2,3,4,5,6,7	100 %	97.2 %	9.5 %
2	8,9,10,11,12,13,14	100 %	100 %	42.9 %

4.2. Optimal assignment of plant operators

In this section, all operators were optimally assigned by using the proposed method. The optimal operator assignment problem was solved using a standard NIP solver, where $A_{a,min}$ was set as 100% for two shifts. Twenty-four patterns of operator assignment were found as optimal solution, and the optimal value of the objective function A_e was 33.3%. Table 6 presents one of the optimal operator assignments and abilities of shifts 1' and 2'. By optimally assigning operators into shifts, the ability of shift 1 in an emergency was drastically improved from 9.5% to 33.3% without sacrificing anything from shift 2. The simulation results proved that we can improve shift's abilities by appropriately assigning operators to each shift.

Table 6 Example of optimal assignment of plant operators

Shift no.	Operator's no.	A_n	A_a	A_e
1'	2,3,5,8,9,12,14	100 %	100 %	33.3 %
2'	1,4,6,7,10,11,13	100 %	100 %	47.6 %

The optimal solution of the operator assignment problem shows us the limitations of operating in shifts when a plant experiences abnormal conditions or in emergency. The evaluation results are helpful for abnormal situation management and for operator training. For example, when a plant condition moves outside of operable conditions of the shift in charge, the shift must ask for assistance to continue operating or shut down the plant safely. It is also worth planning training programs to improve bottlenecks that are found by the optimization.

5. Conclusion

We formulated an optimal operator assignment problem as the mathematical programming problem on the basis of a shift's ability evaluation for industrial chemical plants. The proposed approach is very useful for improving a shift's abilities under abnormal and emergency situations.

The objective function and constraints in the optimization problem can be easily modified according to the characteristics of individual chemical plants.

References

Ikegami, A., Niwa, A., 2003, A Sub-problem-centric Model and Approach to the Nurse Scheduling Problem, *Mathematical Programming*, 97, 517-541.

Noda, M., Kuratsune, K., and Nishitani, H., 2008, Optimal Assignment of Shift Operators for Safety Operation and Skill Advancement, *Proceedings of FOCAPO 2008*, 311-314.

Takeda, M., 2002, Activity for Environment and Safety in the Iwakuni – Ohtake Works, *Human Factors in Japan*, 7(1), 54-64. (in Japanese)

2073

An Eco-Efficiency Study for a WEEE Recovery Network: The Portuguese Case

Maria Isabel Gomes-Salema[a], Ana Barbosa-Povoa[b] and Augusto Q. Novais[c]

[a]*Centro de Matematica e Aplicacoes, FCT – UNL , Monte da Caparica, 2829-516
Caparica, Portugal. e-mail: mirg@fct.unl.pt*
[b]*CEG-IST, Av. Rovisco Pais, 1049-001 Lisboa, Portugal.*
[c]*Dep. Modelacao e Simulacao, INETI, Est. do Paço do Lumiar, 1649-038 Lisboa,
Portugal.*

Abstract

The rapid growth of electric and electronic equipment waste (WEEE) transformed this
waste stream into a worldwide problem. The Directive 2002/96/EC on electrical and
electronic waste aims at the reduction of the environmental impact of WEEE,
encouraging end-of-life management, eco design, life cycle analyses and extended
producer responsibility. However, this legislation may not produce the results the
legislator aimed for. In this work we analyse the environmental impact of a WEEE
recovery network in the Portuguese context. With this aim, a model, previously
developed by the authors for the optimal design of this network using economic
indicators (Salema, 2007), was now adapted to design the network subjected to the
minimization of environmental performance indices.
The original mathematical formulation was found to be flexible and easily adapted to
the two types of indices and the major differences between the optimal network
configurations obtained, were identified and discussed.

Keywords: Recovery network, environmental impact, optimization.

1. Introduction

The rapid growth of electric and electronic equipment waste (WEEE) transformed this
waste into a worldwide problem. The European Union estimates a growth of 3 to 5%
per year for the WEEE, a figure three times greater than for general waste. Around 90%
of WEEE is going to landfills. In order to deal with the problem the Directive
2002/96/EC on electrical and electronic waste was defined that aims at the reduction of
the environmental impact of WEEE, encouraging end-of-life management, eco design,
life cycle analyses and extended producer responsibility. Under this directive, producers
are responsible not only for new products placed on the market, but also for those
equipments sold before this directive. This represents a new driving force that will
compel producers to support collecting and recycling costs of their products.
In Portugal, a non-profit organization (Amb3e) was created by a group of 57 equipment
producers, whose mission is to design and manage the integrated system for the
recovery of WEEE.
Within this organization producers were able to deal with this challenge in a more
efficient way since it allows for economies of scale, reducing the usual uncertainty
associated to the quantity and quality of end-of-life equipments.

Although these recovery systems aimed at the reduction of the environmental impact, the collection of products to reduce the amount of hazardous materials in landfills and the recovery of valuable disposed materials, may still lead to a potential increase of global warming, since products have to be transported to the recovery/treatment facilities.

Several tools have been developed to evaluate the environmental impacts of products and processes. Life-Cycle Assessment is one of them, which evaluates the emissions generation and material consumption during the entire product life. The   eco-indicator'99 (Goedkoop et al. 2000) aggregates into one single index the information on the multiple aspects involved in an environmental impact assessment, taken from the life cycle analysis methodology Also carbon footprint measures in terms of greenhouse gas emissions are very useful, amongst others. More recently, the eco-efficiency concept has been gaining importance in the business community since it combines the economical and environmental objectives. All these tools provide environmental indicators that can be very useful in the design of environmental friendly recovery networks. Such tools have not yet been explored in what concerns the definition and operability of supply chains. Traditionally, the scientific community has been studying supply chain designs through the optimization of economical indicators. This situation, however, is now changing, since current environmental problems ask for new designs that also account for environmental impacts of transportation, production, water consumption and raw-material depletion, recycling, among others.

Few works have been published that evaluate the environmental impact of these structures. Some have nonetheless been published that consider a single type of electric and electronic equipment (Krikke et al. 2003). Neto et al. (2008) proposed a methodology to assess the eco-efficiency in logistics networks and developed a multi-objective model, which was applied in Germany to a recovery network for entertainment equipments, such as TV sets and Wi-Fi equipments.

In this work we analyse the environmental impact of a WEEE recovery network in the Portuguese context. A previous work from the authors on the design of this network using economic indicators (Salema, 2007) is now extended to treat this new problem. The two network structures are compared and the most important aspects affecting each performance indicator highlighted.

This paper is organized as follows. In the next section the context for the Portuguese network is presented. Next, in section 3, the methodology for the impact assessment and a detailed description of the model are given. This is followed by an analysis on the results. Lastly, some final remarks will be drawn.

2. The Portuguese recovery network details

The Portuguese recovery network for WEEE has a two echelon network structure formed by the WEEE sources (that given the strategic nature of this problem are the 278 municipalities of the mainland Portuguese territory), the collection centres that will collect and sort products according to five pre-defined categories, and the recycling facilities already operating in Portugal.

The recycling facilities are located at Tondela and Setúbal. The latter only recycles lighting equipment, which leaves the former with the whole of the remainder. Amb3e estimates that lighting equipment represents just about 1.5% of the total volume to be collected. Tondela facility will therefore process the remaining 98.5%. This will have a great impact on the network, as it will be shown in section 4.

Amb3e needed to perform contractual agreements with companies and organizations that will act as collection centres. Therefore, the major goal is to determine which municipalities are the best locations for collection centres. Together with the locations, some planning information is also needed to better operate this network.

For the year of 2006, Amb3e set the target of recovering 3 kg per inhabitant per year. This objective would allow them to achieve 75% of the legal target imposed by the European directive. This implied that about 30000 tonnes had to be collected and at least 80% of it should be recovered. A maximum distance was imposed so that each Portuguese municipality had a collection centre within a 65 km radius.

In this work, the primary objective is to choose the best locations for the collection centres. In addition, the optimal planning of the network, for the year 2006, will also be determined. This will be possible by modelling a time horizon that is divided in planning units. For the case below the planning units is the trimester.

3. An Eco-efficient Network Model

3.1. Assessing the environmental impact

Several methodologies have been developed over the last two decades to assess environmental impacts of products and services. Life cycle assessment (LCA) is the most applied methodology. It analyses the entire product life, starting at the extraction of raw-materials down to its use by the final customer, and measures each and every impact it has on the environment (Daniel et al., 2004). This involves a very complex process and in order to measure it, some indicators have been developed: eco-indicator99 that measures impacts using a specific unit called points, the carbon footprint which measures impacts in terms of CO_2 emissions, the eco-cost2007 that converts impacts into Euros, etc.

The EU has done an extensive analysis of the impacts that the EU directive on WEEE had on the Member States (Huisman et al., 2008). This report evaluates environmental impacts with several indicators and methodologies. One of the greatest challenges of any work related to the recovery of WEEE is the diversity of products that fit this designation. Products can go from a small hairdryer or a kid's toy, to a refrigerator, LCD television or even an automatic dispenser machine. Therefore it is very difficult to provide an accurate measurement of the environment impact of each of the 10 categories established by the legislation. In Huisman et al. (2008) subcategories were created and impacts were assessed considering average values within each subcategory. Each of these groups of products is then characterized in terms of average weight and by several environmental impact categories.

For the present work, a higher level of product aggregation is needed since Amb3e only considers five groups of products: large household appliances (C1), cooling and freezing equipment (C2), small household appliances (C3), monitors and televisions (C4), lighting equipment (C5).

In terms of environmental measures, we considered four major impact sources: **the impact of establishing the collection centres**: this impact reflects the construction of a facility similar to a warehouse; no assessment was made regarding the impact of the land use; **the transportation impact between WEEE sources and collection centres** (by km and tonne): this value was based on the impact values provided by TU Delft University (Eco-costs); **transportation impact between collection centres and recyclers** (by km, product and tonne): this value was based on the values of transportation given in the EU report; in categories with more than one subcategory, the average value is weighted by each product own weight; **the recovery impact**: this

reflects the gain in terms of environment for the recycling materials, therefore this is a negative value; and **the disposal impact**: that reflects the impact of sending to proper disposal products that could not be recovered; this operation does not have to be bad for the environment since some products can be incinerated, producing energy.

No impact value was given to products kept in stock. It was assumed that products were well conditioned and, since at the end of the planning period, all collect products have to be recycled, these stocks do not represent any harm to the environment. Nonetheless and since legal targets have to be meet, it was assumed that in the last period of the planning horizon.

3.2. The model

A model, previously developed by the authors (Salema, 2007), was adapted to design the network according to the environmental performance measures. The same model is used to minimize impacts and (but not simultaneously) to minimize costs. Therefore, the parameters defined for the objective function have in each case a different meaning, either impacts or costs. For reasons of confidentiality the real units are omitted and values presented are expressed in terms of monetary units (m.u.) or impact units (i.u.). Due to lack of space, only a general description of the model is next given. The model can be stated as: Given: WEEE volume to be collected at each source; recovery target set by legislation; sorting criteria to be performed at centres; initial stock levels at facilities; maximum storage capacity; maximum and minimum processing capacities; upper and lower bounds for flows; unit impact/ compensation fee given to recyclers and sorting centres; unit transportation impact/cost; unit processing and storage impact/cost. Determine, locations of sorting centres; flows amounts between sources of WEEE and sorting centres and between these and the recycling facilities; storage volumes at sorting centres and recycling facilities; processed and disposed volumes by recycling facilities, so as to minimize the total environmental impact or total cost of the network. In terms of constraints, the model is characterised by nine types of constraints related to: material balances; collection satisfaction; target levels assuring that the legal target is met; disposal, since it is estimated that 5% of the collected products are not proper to recycle; maximum stock levels; maximum and minimum values for flows; and finally sorting centres capacities limitations. In terms of variables, the model has three types of continuous variables (flow amounts, stock levels and non-satisfied return), and one type of binary (choice of entities to integrate the network). For further details regarding the model refer to (Salema 2007).

4. Results and discussion

The model described above was run in its original and modified versions: the former considering the economic parameters (cost model) and the latter the environmental impacts (impact model). All results presented were computed on a Pentium 1.7 GHz, using GAMS/CPLEX (built 22.9). First, the results obtained for the impact model will be presented, and then the results of both models will be discussed.

4.1. The environmental impact network

The optimal location of collection centres for the environmental impact model is depicted in Figure 1. The network is formed by 110 centres spread all over the country. The total impact of this network is $-3312*10^3$ i.u. Note this is a negative value, which means that the recovery of products is environmentally friendly. In a "doing nothing" context, the disposal of products in the municipal waste system in a uncontrolled landfill without energy recovery, would have an positive impact of about $371*10^6$ i.u. (this value was calculated using the data provided at Huissman et al (2008)).

Looking closer to the positive impacts shown in Figure 2, one sees that the transportation between collection centres and recycling facilities (flow F2) is responsible for 97% of the total positive impact ($157*10^3$ i.u.).

Comparing flows F1 and F2 (F1 refers to the connection between sources and collection centres), the first conclusion taken it that the location of recycling facilities is of extreme importance and, for this case in particular, at least a second unit should be considered. Although the construction and operation of a recycling unit has a positive contribution in terms of impacts, the decrease in the transportation impact most certainly will overshadow this negative effect.

This model produces a large amount of information regarding network planning: recycling, transportation, disposal, and storage. Due to the lack of space, only a few examples will be shown.

The first is related to the recycling plan, Figure 3. This figure illustrates very clearly the difference between both recycling facilities in terms of processed volume. Another important aspect are the disposal amounts. As mentioned above it is estimated that at least 5% of collected products will not be suitable for recycling, and should be properly disposed in landfills and, when possible, energy should be recovered. The optimal disposal volume is exactly 5%. The reason is that recycling activities reduce the environmental impact of the network. It will be shown below that the opposite scenario characterizes the cost model solution.

Figure 1: Environmental impact recovery network

4.2. Comparison between cost and environmental networks

Using the same model structure, the formulation mentioned above was used to determine the optimal solution in terms of economic objectives. The environmental impact of the optimal solution for the cost model was computed, as well as the cost related to the optimal solution of the impact model. All values are shown in Table 1.

This table shows some interesting aspects. The transportation costs between WEEE sources and collection centres are much higher in the impact solution. This is explained by the number of centres which is much smaller: in the impact solution there are 110 centres, while in the cost solution these are 169. Therefore, for the latter more travelling is needed.
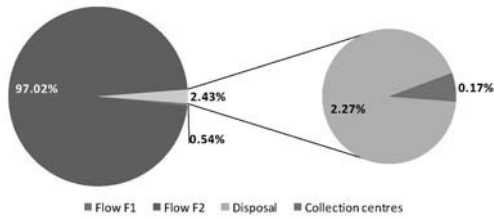
Figure 2: Weight of the environmental parameters in the total impact value.

With regard to transportation between collection centres and recycling facilities, environmental impacts show a larger difference than costs. More centres reduce the total travelling. Disposal costs present also significant differences. Note that these values (both for environmental impacts and costs) are four times higher in the cost model. The legal target imposes that 80% of collected products to be recycled. In the solution of the cost model, only 80% of products

Figure 3: Recycling and disposal plans for the entire time horizon.

are recycled while in the impact model this percentage is 95%. The reason is that in the cost model, the recycling represents a cost and in the impact model this operation has a negative impact. Stocks only appear in the impact model. As mentioned above, no impact was assessed for products kept in stock. Therefore, the impact solution has high levels of stocks. These are, however, accounted for when costs are calculated.

Table 1: Environmental impacts and costs in the optimal values of both models.

	Impacts (u.i.)		Costs (m.u.)	
	Impact model	Cost model	Impact model	Cost model
Recycling	-3 463 106	-2 916 300	9 169 466	7 721 656
Collection	253	253	782704	782 704
Transp. F1	820	58	1 241 304	88 483
Transp. F2	146 268	174 992	5 176 062	6 176 314
Disposal	3 415	13 659	73 840	295 360
Stocks	0	0	1 062 436	0
Total	-3 312 350	-2 727 338	17 505 812	15 064 517

5. Final remarks

In this work, a recovery network for WEEE recovery in the Portuguese context is designed considering environmental performance metrics. The optimal location for collection centres is presented and some planning details discussed. With the original model, the optimal network was also designed and planned considering this time an economic objective function. Both solutions were presented and compared.

As mentioned above, it is very hard to estimate accurately the environmental burden of each product category. Therefore, one way to overcome this limitation is to perform similar studies using different methodologies for impact assessment. Further work is being done to develop a methodology for assessing the environmental impact of global recovery networks.

References

Daniel, S. E., et al. (2004). *Ecological Indicators*, 4(2), 125-138.
Eco-costs, http://www.ecocostsvalue.com/httpdocs/content/html/5-data.html.
Neto, J. Q.F., et al. (2008). *IJPE*, 111(2), 195-208.
Goedkoop, M., et al. (2000). PRé - product ecology consultants.
Huisman, J., et al. (2008). United Nations University, Bonn, Germany.
Krikke, H., et al. (2003). *IJPR*, 41(16), 3689-3719.
Salema, M. I. (2007). PhD thesys, Instituto Superior Técnico, UTL, Lisbon.

Valuation of Clean Technology Projects: An Application of Real Options Theory

Marcelo C. M. de Souza[a], Cristiano H. O. Fontes,[a] Silvio A.B. Vieira de Melo,[a]
Antonio Francisco A. Silva Junior[a]

[a] *Programa de Engenharia Industrial - PEI, Escola Politécnica, Universidade Federal da Bahia – UFBA, R Aristides Novis, n. 2, Federação, Salvador - BA, 40210-630, Brasil*

Abstract

The valuation of investment in technologies that minimize the environmental impacts is not an easy task. Investment and environmental damages are partially or totally irreversible, the forecast period for the analysis of such investment is usually long and there are economic and environmental uncertainties which are still enhanced by the fact that environmental damages is not a linear function of the impacts. A better understanding of the behavior of these uncertainties and of the strategies to incorporate these uncertainties in the design valuation would be useful in the decision making associated to the investment with relevant environmental impacts. This work presents an extensive review about the application of the Real Options Analysis in the investment decision associated to the environmental problems and discusses the potential of this technique in the investment analysis of these cases. The aspects discussed denote that the approach of Real Options Analysis is better than the traditional methods and possibly is capable of predicting the benefit obtained with the reduction of the environmental risks provided by the use of clean technology.

Keywords: Real Options Analysis, Clean Technology, Environmental Evaluation.

1. Introduction

The environmental sustainability debate, focusing on the ability to combine economic growth with the environment preservation, has been broad and deep. According to Jeffrey Sachs (2008), until now, just one thing is true: the current trajectory of human activity is not sustainable. If the humanity maintains the same current technologies considering the scale up due to the fast economic growth of big population centers such as China and India, the environment and humanity's welfare will collapse.

The traditional production process has caused social and environmental problems with increasing impact potential (SEIFFERT, 2007). Only more recently, the practice of pollution prevention and clean production strategies has begun to be adopted in order to reconcile economic performance together with improvements in the environmental performance (KIPERSTOK, 2008). In most cases the adoption of sustainable practices have been employed due to the pressure imposed by the society and many investors still consider the sustainable development a necessary (but not desirable) strategy which involves the adoption of a series of procedures, regulations and onerous responsibilities. They do not consider sustainable practices as a strategic business opportunity (HART and MILSTEIN, 2004).

Generally investor's motivation is just the increase of wealth. The adoption of strategies based on clean technologies can imply more investment when compared with traditional processes, and can also affect the future cash flows and the risk of business through changes in costs and revenues and the reduction of environmental risk. It is usually difficult to predict the real benefits of these practices for the company. Therefore, the environmental risks, usually difficult to quantify, are not considered in decision-making process (BEKEFI and EPSTEIN, 2007).

Traditional investment analysis methods use an approach based on discounted cash flow. This is questioned by professionals and academics due to its inability to capture the administration's flexibility to adapt and revise subsequent decisions caused by market changes, being incapable to evaluate all value sources of an investment (TRIGEORGIS, 1993, SIEGEL et al. 1987, TRIGEORGIS, 2007). In these methods the uncertainty related to the project and the reactions to the market conditions changes are treated superficially, changing the rates to update the risk (MORCK et al. 1989). Thus, the higher the uncertainty and flexibility of management to react to market changes, the more undervalued will be the project value using traditional methods of evaluation (COPELAND and ANTIKAROV, 2001).

More recently, a new framework for investments evaluation, called as Real Options Analysis (ROA), has treated uncertainties in investments design. The purpose of this article is to review the state of the art and evaluate the application of the Real Options Analysis in the investment decision associated to environmental problems and clean technology projects. Section 2 presents the concept of ROA and presents some arguments that justify its use in environmental analysis. Section 3 presents a review about ROA applications in works about environmental valuation and clean technology and also shows why the ROA application in environmental analysis is still incipient. Finally, the last section discusses the basic aspects that suggest the adoption of ROA in the valuation of clean technology projects.

2. Why use Real Options in the Environmental Investment Analysis?

Traditional methods of investment analysis based on discounted projected cash flows have some limitations with regard to the capture of the management flexibility. Initially these methods consider that the investment is reversible. Even when this is not true, if the firm does not undertake the investment now, it will not be able to in the future (DIXIT and PINDYCK, 1994). In addition the Net Present Value rule assumes that the decision initially taken can not change in the future.

The Real Options Analysis does not replace the Net Present Value (NPV), but rather represent a complement of this technique. The ROA can capture the value of the project flexibility resulting in a new NPV which some authors call the NPV expanded. These management flexibilities are basically related to the strategic positioning of the company. So, the analysis of real options can provide a connection between the strategic planning and the finance of the company. The possibility of a new entrant in the market, the possibility of a delay in the implantation of the project due to the wait for more information that would reduce the uncertainty, expand, contract or even abandon the business are all management flexibilities of the project. In fact, if a project has

flexibility options, the investment opportunity value is equal to the Net Present Value plus the value of the options.

The investment in projects associated with clean technologies is difficult to evaluate because it represents a direct cost for the investor and the return is not direct. The benefits provided by clean technologies comprise the reduction of pollution levels, the reduction of the costs for its treatment, the compliance to the environmental laws, the reduction of probability of environmental damages, the reduction of shutdown events and reputation risks that could imply in loss of market share and even the solvency of the company among others. But it is not so clear that the benefits will arise in the next production cycle. The environmental damage risk or the pollution level represents an externality. The externality is considered in some way a possible contingency in the project. An environmental contingency can be seen as an obligation to be exercised or not. If it is possible to calculate the price of this option it is possible to evaluate the project considering its externality.

Real Options Analysis (ROA) can provide a consistent method for valuation of projects with high uncertainties through the analogy between the financial options and the options for business management. It uses financial options pricing methods to evaluate this latter and has been presented as an alternative, which leads to better results in the evaluation of investment, especially in environments of high uncertainty (DIXIT and PINDYCK, 1995) (COPELAND and ANTIKAROV, 2001).

Considering that the effects of environmental impacts may be irreversible and that the value of environment preservation for future generations is uncertain, the benefit in the environment protection today must include a value of option. When this option is exercise, the managerial flexibility is lost. The social costs deriving from the anticipated exercise of this option can be large and should be measured.

3. ROA applications in Environmental Valuation.

The first works of environmental preservation presented in the literature that established the concept of irreversibility are due to Arrow and Fisher (1974) and Henry (1974). The option value concept in the economic environmental approach was developed and explained by Fisher and Hanemann (1987) and Hanemann (1989).

Recently some applications of real options theory in issues about environmental policy have been suggested, often to solve optimal timing problems due to the uncertainties and the irreversibility associated to the environmental degradation, its prevention and its economic consequences (Pindyck, 2000, 2002). Pindyck (2002) shows that the traditional approach using the net present value of the costs and benefits flows as a rule of decision, ignore three important features of the most environmental problems: i) the uncertainty of the future costs and benefits, ii) the irreversibility associated to the environmental policy, and iii) the possibility to delay some action and to wait for more information. He used a two-period model to illustrate the optimal timing problem and its implications on the environmental policy.

Dotsis et al (2005) extend the analysis of Pindyck (2000) allowing the state variables to follow jump diffusion processes. They analyzed the features of optimal timing decisions

related to the adoption of an environmental policy when there is a possibility of abnormal changes either in the evolution of the stock of pollutant, or in the future costs and benefits for environmental damage.

Lin, Ko, Yeh (2007) propose a continuous time model using the real option approach, (extension of Pindyck model), capable to determine when and how much greenhouse gas emissions should be reduced. The model uses a geometric Brownian motion to describe the economics and ecological uncertainties and also considers the existence of correlation between these variables.

Martins and Melo (2007) use Real Options Theory to analyze if a larger urban mangroves area must be used for the construction of a highway. They also use a continuous time model that considers a geometric Brownian motion to describe the benefits flow provided by the mangroves area.

There are few works using Real Options Analysis approach to evaluate environmental questions associated to private investment analysis. Sarkis and Tamrkin (2005) apply real options analysis in a real case study (*British Petroleum – Amoco*) for valuation one specific project with carbon credit generation. Espinoza and Luccioni (2007) present a model derived from an option to estimate the value of brownfields considering the environmental and market risks.

In another recently study, Pindyck (2007) shows the importance of uncertainties in the evaluation of environmental policies and attests that traditional models of cost benefits analysis are not suitable for evaluation of these policies because these have some limitations, namely, the quantification of the reduction in the environmental damage, the benefit from this reduction are unknown, the physical and ecological impacts are uncertain and there is relevant uncertainty related to the costs and benefits associated to the policy that will be employed; The discount rate to be used is not know exactly. Uncertainty about costs, benefits and discount rates are common in investment analysis, but, according to Pindyck (2007), three additional difficulties arise:

1. The environmental cost and benefit functions tend to be highly nonlinear and expected values are difficult to be used in this case.
2. Environmental policies usually include significant irreversibility, and they sometimes interact in complex way with uncertainty. There are sunk costs for the society (such as investment or as a discrete flow of expenditure). Furthermore, environmental damage is often partially or completely irreversible..
3. Environmental policies often comprise long time horizons. This increases the uncertainty associated to the cost, benefit and discount rate.

An irreversible investment opportunity is similar to a financial Call Option. A Call Option gives to the holder the right to acquire an active at a fixed price. The exercise of this option is irreversible. If the active value increases, the net return of the investment increases, however if the active value falls the investor does not exercise the option (to invest) and loses it was paid to have the right to invest.

A literature search was carried out in the database *Engineering Village* and the results are shown in Table 1 in number of publications identified. Combinations for "clean technology" or "clean production" and key words which mean processes of economic or financial assessment, such as: "Real Option" Investment Decision"," Investment Analysis" were not found.

Keywords — Subject, Title and Abstract
Período — 1999 - 2008
All Languages

And		"Real Option*"	"Investment decision"	"Investment Analysis"	"Clean* Technology"	"Clean* Production"	"Enviromental Valuation"	"Sustainable Development"	"Enviromental Economic*"	"cost benefit analysis"
"Real Option*"	211		29	5	0	0	0	5	0	11
"Investment decision"	400			10	0	0	0	16	0	14
"Investment Analysis"	107				0	0	0	3	0	8
"Clean* Technology"	147					5	0	18	0	2
"Clean* Production"	94						0	14	0	2
"Enviromental Valuation"	47							1	1	7
"Sunstainable Development"	17.815								70	217
"Enviromental Economic*"	216									12
"cost benefit analysis"	6.820									

Table 1 - Results of literature search

"Sustainable development" and "Cost benefit analysis" produced the largest number of 217 results with great concentration in the years 2003 to 2005, still high in 2006 and 2007, confirming that most of the analysis of investment models that address issues of sustainability are still based on the method of cost-benefit analysis. Only 5 works are directly related to the ROA application in sustainable development.

First of all, the environmental issues are relatively recently. After then, there are many difficulties in modeling. There are not many data to assess the stochastic parameters, the process of learning and emergence of new information and new technologies is very dynamic. The analysis tools based on ROA is much more complex than the traditional strategies and sometimes is difficult for the investor understanding. The stochastic behavior is difficult to modeling and often involves jump diffusions and optimal stopping problems solution.

4. Conclusion.

Based on the literature review and on the intrinsic features associated to the evaluation of environmental projects, this paper suggests that the traditional cost-benefits analysis often employed in these tasks have some misleading. The main difficulties presented in these problems, namely, total or partially irreversibility over investment, cost and benefit, uncertainty over the future costs and benefit and managerial flexibility to change the choices in function of market and technological changes are the same as those in which the ROA is based on. Therefore, Real Option can represent a potential tool to support the decision making in environmental projects, specifically related to clean technologies. The few works that present ROA application in this kind of problem attest the complexity of the modeling considering all the tools provided by this method.

References

K. J. Arrow and A.C. Fisher (1974) Environmental Preservation, Uncertainty, and Irreversibility. The Quartely Journal of Economics, Vol. 88, No. 2 p. 312-319.

T. Bekefi and M. Epstein (2007) The whole picture CMA Management; 80, 8 pg.23-26.

T.E. Copeland and V. Antikarov (2001) Opções Reais: um novo paradigma para reinventar a avaliação de investimentos. Tradução de Maria José Cyhlar. – Rio de Janeiro: Campus.

A.K. Dixit and R.S. pindyck (1994) Investment Under Uncertainty. Princeton, New Jersey: Princeton University Press.

A. K. Dixit and R.S. Pindyck (1995) The Options Approach to Capital Investment. Harvard Business Review, May.

G. Dotsis, V. Makropoulou and D. Psychogios (2005) The Timing of Environmental Policies in the Presence of Extreme Events. Real Options – Theory meets practice - 9th Annual International Conference. Paris – France, June 22-25. Available: http://www.realoptions.org/abstracts/abstracts05.html.Accessed in 29/10/2007.

R.D. Espinoza and L.X. Luccioni (2007) An approximate solution for perpetual American option with time to build: the value of environmental remediation investment projects. International Journal of Business. Volume 12 number 3 summer. Available: http://www.craig.csufresno.edu/IJB/Volumes.htm. Accessed in 24/08/2007.

A.C. Fisher and W.M. Hanemann (1987) Quasi-option Value: Some misconceptions Dispelled, Journal of Environmental Economics and Management, 22, pp. 197-204.

W.M. Hanemann. (1989) Information and the Concept of Option Value, Journal of Enviromental Economics and Management 16, pp. 23-37.

S. L. Hart e M. B. Milstein (2004) Criando Valor sustentável. Revista de Administração de Empresas – FGV, Vol 3 n° 2 maio/jul.

C. Henry (1974) Investment Decisions Under Uncertainty: The "Irreversibility Effect". The American Economic Review, Vol 64. No. 6 Dec. Pp. 1006.1012.

A. Kiperstok (2008) O papel da Universidade e da Rede Teclim na Introdução de Práticas de Produção Limpa na Bahia. In KIPRSTOK, A. (org.) Prata da Casa: construindo produção limpa na Bahia. – Salvador, Rede de Tecnologias Limpas da Bahia – Teclim/Universidade Federal da Bahia.

T. Lin, C. Ko, H. Yeh. (2007) Applying real options in investment decisions relating to environmental pollution. Energy Policy 35 p. 2126 – 2432.

G. M. Martins & A.S.S.A.Melo (2007) O Valor Da Preservação Do Parque Dos Manguezais Em Recife-Pe: Uma Utilização Do Método De Opções Reais, Anais do XXXV Encontro Nacional de Economia [Proceedings of the 35th Brazilian Economics Meeting] 146, ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics]. Avalilable: http://ideas.repec.org/p/anp/en2007/146.html. Accessed in 05/01/2009.

R. Morck, E. Schwartz and D.Stangeland (1989) The valuation of Forestry Resources under Stochastic Prices and Inventories. Journal of Finance and Quantitative analysis. Vol. 24, N° 4, PP.473-487.

R. S. Pindyck (2000) Irreversibilities and the timing of environmental policy. Resource and Energy Economics. Vol. 22 PP. 233-259.

_____ (2002) Optimal timing problems in environmental economics. Journal of Economic Dynamics & Control. Vol. 26 PP. 1677-1697.

_____(2007) Uncertainty in Environmental Economics. Review of Environmental Economics and Policy, volume 1, issue 1, PP. 45-63, winter.

J. Sachs (2008) Common Wealth: Economics for a Crowded Planet. Great Britain, Allen Lane.

J. Sarkis and M. Tamarkin (2005) Real Options Analysis for "Green Trading": The Case of Greenhouse Gases. The Engineering Economist, 50 p. 273 – 294.

M. E.B. Seiffert (2007) Gestão Ambiental: instrumentos, esferas de ação e educação ambiental – São Paulo: Atlas.

D.R. Siegel, J.L. Smith, J.L. Paddock. (1987) Valuing Offshore Properties With Option Pricing Models. Midland Corporate Finance Journal, Vol. 5, Spring. PP. 22-30.

L. Trigeorgis (1993) The nature of option interactions and the valuation of investments with multiple real options. Journal of Financial and Quantitative Analysis, p. 1-20, mar.

_____ (2007) Opções Reais e interações com a flexibilidade financeira. Revista de Administração de Empresas – FGV, Vol 47 n° 3 jul/set.

Crude Oil Transshipment Using Floating, Storage, and Offloading Platforms (FSOPs)

Sangeeta Balram and I.A. Karimi

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576

Abstract

Floating Storage and Offloading Platforms (FSOPs) have been around for forty years now and are attracting attention due to increasing offshore fields and facilities for oil and gas. Although in a conceptual stage, this technology can enable more efficient crude delivery to various demand centers onshore. The concept of FSOP has been prevalent for some time now in terms of its design and sustenance offshore. Although, the use of FSOPs for crude transshipment has been existential in the logistics industry, yet no work has so far addressed the management of such units for crude oil handling and transportation. To address this research gap, we combine a mix of techniques used by Cheng and Karimi (2006) **as well as** Reddy and Karimi (2004) **specifically aiming at building continuous time models.**

Keywords: FSOP, MILP, VLCCs/ ULCCs

1. Introduction

World crude oil demand is expected to increase to 90 million barrels per day by 2010, a growth of around 43% in a span of 20 years (1990-2010). In order to meet this heavy demand, world offshore oil production has risen three times to 53 million barrels per day in a span of 10 years whereas the world offshore gas production has more than doubled during the same span (PennWell Corporation, 2009). Offshore oil production now accounts for one third of the total world production and is growing at a much faster rate than the onshore sector. Over the past many years, Oil and Gas industry has thrived heavily on offshore drilling and storage units for ease of transportation of crude oil onshore. A technology, which has already been proposed in industrial literature, is the use of Floating Production & Storage Platforms (FPSOP) offshore. Although in a conceptual stage, this technology can enable more efficient crude delivery to various demand centers onshore. Crude oil transportation and delivery is a major trade activity in the oil and gas industry, largely done using Very Large Crude Carriers (VLCCs) or Ultra Large Crude Carriers (ULCCs). So far, crude oil transshipment by tanker lightering has been in practice in many fields. This practice is however, limited by the inability of VLCCs or ULCCs in entering ports with shallow draughts. The FPSOP can not only act as an intermediary between the oil market and refineries, in coordinating oil trade between the two, but can also act as an intermediate storage unit for crude parcels during transshipment, reducing onshore inventory levels. FPSOPs especially facilitate crude storage and transportation in remote or deep waters where pipeline infrastructure is not cost effective. Pipeline infrastructure in many fields may increase disproportionately with water depths. In smaller or marginal fields, oil and gas may be

produced for a relatively short duration of time. Laying fixed bed platforms or pipelines can prove extremely capital intensive in such fields. FPSOPs are considered more economical systems for oil and gas drilling and storage especially in the marginal fields. Ierapetritou (2003) divided their continuous time formulation into a number of subproblems, the first comprising crude oil operations, second dealing with refinery processes and intermediate tanks, and the third related to finished products and blending operations. Jetlund and Karimi (2003) worked on efficient routing and scheduling of chemical tankers for improving logistics in global chemical supply chains. They developed MILP formulation for a single ship and extended it to maximize profit for a fleet of multi-parcel tankers engaged in shipping liquid bulk chemicals. Reddy and Karimi (2004) built discrete and continuous time models for optimizing refinery operations while taking different tanks as intermediate storage units before transferring crude oil to several Crude Distillation Units (CDUs).

2. Problem Definition

In this work, we consider our FPSOP as an FSOP owned and managed by a single centralized authority. The system consists of several vessels like VLCCs/ULCCs ($v = 1, 2, \ldots, V$) connecting through an SBM pipeline to the FSOP. A number of small Service Ships ($n = 1, 2, \ldots, N$) (SSs, Cheng & Karimi, 2007a, b) or pipelines are employed for offshore delivery to a set of refineries ($r = 1, 2, \ldots, R$). The FSOP consisting of tanks ($s = 1, 2, \ldots, S$) of different capacities acts as a central procurement unit between the oil market and refineries making the main purchasing decision of crude from the oil market. The refineries work in collaboration with each other for mutual benefit (Collaborative Logistics). The FSOP acts as a trans-shipment hub for crude supply to the refineries based on orders ($o = 1, 2, \ldots, O$) from the refineries and charters out vessels, bringing a variety of crude parcels ($p = 1, 2, \ldots, P$) into the tanks. We assume that each vessel scheduled to supply crude to FSOP has an estimated Earliest Docking Time (EDT_v). Furthermore, since only one vessel can dock at any time, we assume that no two EDTs are identical. Thus, we can sequence the EDTs in the chronologically increasing order. A vessel is stipulated to depart anytime at or before its Stipulated Time of Departure (STD_v). The tanks serve as storage, blending and offloading units. The refineries' orders contain multiple crude products ($c = 1, 2, \ldots, C$) of fixed composition according to refineries' specifications. Each order is related to a delivery time window $[DD_o^L, DD_o^U]$. Refineries most commonly use 15 crude properties m which form a basis for specifying the crude quality (Li et al., 2007). These blending indices are either volume based or weight based.

Given:
- A scheduling horizon $[0, H]$.
- V vessels, their estimated arrival and departure times.
- P parcels, their total volumes and flow rates.
- S tanks, their capacities, outgoing flow rates, initial inventories and limits on their hold ups.
- C crudes, their property indices and specific limits on their property indices.
- O orders, their constituent crudes and required amounts.
- Economic data such as demurrage or delay costs (vessels and orders), inventory costs, changeover costs, set up costs (tanks).

Determine:
- A detailed schedule for parcel unloading from vessels and their amounts.
- A detailed schedule for order delivery from tanks and their amounts.

Operating policies:
- Owing to SBM pipeline, only one vessel can dock at any time and atmost one parcel can be unloaded from it at a time.
- A tank cannot receive and deliver simultaneously.

Assuming:
- There is no hold-up in the SBM pipeline connecting the vessel to the FSOP.
- Changeover times and set up times in tanks are negligible.
- All costs are constants and any price volatility with respect to costs is ignored.
- FSOP is physically stable and any stability fluctuations from turbulence in weather conditions are ignored.

3. Mathematical Formulation

Each vessel may carry multiple single-crude parcels. Since the unloading sequence of parcels from each vessel is usually pre-fixed (Reddy et al., 2004), and it takes non-zero time to transfer each parcel, we can also define a fixed sequence for unloading all crude parcels from the vessels. We now define an Earliest Time of Arrival (ETA_p) for each parcel p as the unique time at which the parcel may become available for transfer. We compute $ETA_p = EDT_v + \sum_{p=2}^{P} \tau_{p-1}$ for all $p \in \mathbf{P}_v$ where $\mathbf{P}_v = \{p \mid$ vessel v carries a parcel p of some specific crude}. Note that, the Earliest Time of Arrival (ETA_p) of any parcel p is different from its actual time of transfer, which is decided by the optimizer based on the unloading sequence of parcels. Time representation is of prime importance in any scheduling problem. We propose an asynchronous slot based approach to model time on the SBM pipeline and tank s independently in terms of a series of K chronologically ordered contiguous slots ($k = 0, 1, .., K$) of variable lengths. Slot k ($k = 1, 2, .., K$) on the SBM pipeline starts at $T_{(k-1)}$, ends at T_k and has a length of $[T_k - T_{(k-1)}]$. Similarly, slot k on tank s starts at $T_{s(k-1)}$, ends at T_{sk} and has a length of $[T_{sk} - T_{s(k-1)}]$. Denoting the start of scheduling horizon as the start of slot 1, we have $T_0 = T_{s0} = 0$. By definition;

$$T_k \geq T_{(k-1)} \qquad\qquad T_k \leq H , 1 \leq k \leq K \qquad\qquad (1a)$$

$$T_{sk} \geq T_{s(k-1)} \qquad\qquad T_{sk} \leq H , 1 \leq k \leq K \qquad\qquad (1b)$$

We assume that a real operation starts only at the beginning of a slot and continues during the slot; however it may end at any time within the slot. Thus, the idle time if any, is always towards the end of the slot (Susarla et al, 2008). While parcel transfer is the only operation on the SBM pipeline; tank operations include parcel receipt, order delivery, crude storage, crude changeovers, brine settling as well as crude blending.

3.1 Parcel Transfer Operations

We model the parcel transfer operation on the slots of the SBM pipeline by defining one binary variable x_{pk}.

$$x_{pk} = \begin{cases} 1 & \text{if parcel } p \text{ is transferring during slot } k \\ 0 & \text{otherwise} \end{cases} \qquad 1 \leq p \leq NP, 1 \leq k \leq K$$

As noted earlier, atmost one parcel can transfer through the SBM pipeline during a slot. Therefore,

$$\sum_p x_{pk} \leq 1 \qquad\qquad 1 \leq p \leq NP,\ 1 \leq k \leq K \qquad (2)$$

3.2 Parcel receipt into tanks

Now, if a parcel transfers during a slot of the SBM pipeline, it must also be received during a slot on its associated tank. Let us suppose that the parcel transferring in slot k of SBM pipeline is received by tank s during its slot k'. Now, we can have three possible scenarios such that: $k'<k$, $k'=k$, and $k'>k$. For the first and third scenario, we can simply add extra slots on the tank and SBM pipeline respectively, to make $k'=k$ (Susarla et. al, 2008). Therefore, we demand that if the SBM pipeline transfers a parcel into any tank s, then the slots corresponding to the parcel transfer on the SBM pipeline and parcel receipt on tank s should have the same index. With this explanation, we now denote parcel receipt into tank s by binary variable u_{sk} as;

$$u_{sk} = \begin{cases} 1 & \text{if tank } s \text{ is in a state of receipt during slot } k \\ 0 & \text{otherwise} \end{cases} \qquad 1 \leq s \leq S,\ 1 \leq k \leq K$$

We allow exactly one tank to receive a parcel during a slot. This means;

$$\sum_s u_{sk} = \sum_p x_{pk} \qquad\qquad 1 \leq k \leq K \qquad (3)$$

However, equation (6) does not indicate the transfer of a particular parcel p into its destination tank s during slot k. For this, we introduce a 0-1 continuous variable $xu_{psk} = x_{pk}u_{sk}$;

$$xu_{psk} = \begin{cases} 1 & \text{if parcel } p \text{ transfers into tank } s \text{ during slot } k \\ 0 & \text{otherwise} \end{cases} \qquad s \in \boldsymbol{S}_p,\ 1 \leq k \leq K$$

where $\boldsymbol{S}_p = \{s \mid \text{tank } s \text{ can receive parcel } p\}$
We use McCormick's linearization method to linearize xu_{psk}.

3.3 Material balance for parcels

Let volume of parcel p transferred into tank s during slot k be QI_{psk}. Total amount of each parcel is Q_p. Hence,

$$QI_{psk} \leq Q_p xu_{psk} \qquad\qquad s \in \boldsymbol{S}_p,\ 1 \leq k \leq K \qquad (4a)$$

To ensure that a parcel transfers fully within the scheduling horizon;

$$\sum_k \sum_{s \in S_p} QI_{psk} = Q_p \qquad\qquad 1 \leq k \leq K \qquad (4b)$$

3.4 Timing constraints

Transfer of any parcel p through the SBM pipeline cannot start before its Earliest Time of Arrival (ETA_p). This means;

$$T_{k-1} \geq \sum_p ETA_p xs_{pk} \qquad\qquad 1 \leq k \leq K \qquad (5)$$

Now, the actual amount of parcel transferred into a tank during any slot is limited by its maximum pumping flow rate,

$$\sum_{s \in S_p} QI_{psk} \leq FP_p^U (T_k - T_{k-1}) \qquad\qquad 1 \leq k \leq K \qquad (6)$$

Recall that each vessel has a Stipulated Departure Time (STD_v). If the last parcel from each vessel remains connected to the SBM pipeline beyond this time; demurrage penalty is imposed. Let d_v denote the delay in the departure of vessel v.

$$d_v \geq T_k - STD_v - H(1 - x_{pk}) \qquad\qquad p \in \boldsymbol{LP}_v,\ 1 \leq k \leq K \qquad (7)$$

where $LP_v = \{p \mid \text{parcel } p \text{ is the last parcel from vessel } v\}$

Since a parcel transfers into one of its destination tanks at a time; we need to register the parcel transfer times on tanks as well. For this; we synchronize the start and end times of parcel transfer on SBM pipeline with that on the tank

3.2 Order Delivery Operations

Each tank can deliver several single-crude product orders within the scheduling horizon. We denote delivery of order o during slot k on any tank by a binary variable y_{ok} as follows;

$$y_{ok} = \begin{cases} 1 & \text{if order } o \text{ delivers during slot } k \\ 0 & \text{otherwise} \end{cases} \qquad o \in \boldsymbol{O}_s, \, 1 \le k \le K$$

where $\boldsymbol{O}_s = \{o \mid \text{tank } s \text{ can deliver order } o\}$

Let QO_{sok} be the volume of order o delivered by tank s during slot k. Total required quantity of each order is RQ_o. Now, if a tank does not deliver an order during slot k, then its quantity is zero. Hence,

$$QO_{sok} \le RQ_o y_{ok} \qquad o \in \boldsymbol{O}_s, \, 1 \le k \le K \qquad (8)$$

To ensure full delivery of each order o;

$$\sum_k \sum_s QO_{sok} = RQ_o \qquad o \in \boldsymbol{O}_s, \, 1 \le k \le K \qquad (9)$$

Amount of each order delivered from a tank is limited by maximum delivery rate FS_s^U of each tank such that,

$$\sum_{o \in Os} QO_{sok} \le FS_s^U (T_{sk} - T_{s(k-1)}) \qquad o \in \boldsymbol{O}_s, \, 1 \le k \le K \qquad (10)$$

Each order has a certain delivery time window as mentioned earlier $[DD_o^L, DD_o^U]$. This means a tank s cannot start delivering an order o before DD_o^L;

$$T_{s(k-1)} \ge DD_o^L y_{ok} \qquad o \in \boldsymbol{O}_s, \, 1 \le k \le K \qquad (11)$$

Also, if a tank delivers an order beyond DD_o^U, penalty is imposed under order delivery delay. We consider d_o as the delivery delay for each order.

$$d_o \ge T_{sk} - DD_o^U - H(1 - y_{ok}) \qquad o \in \boldsymbol{O}_s, \, 1 \le k \le K \qquad (12)$$

3.4 Inventory Balance in Tanks

Let I_{sk} denote the current inventory level of tank s during slot k. The initial inventory level of tanks s before time $T_{s0} = 0$ is known to be I_{s0}. Thus, inventory balance in any tank s is given as:

$$I_{sk} = I_{s(k-1)} + \sum_p QI_{psk} - \sum_o QO_{sok} \qquad s \in \boldsymbol{S}_p, o \in \boldsymbol{O}_s, \, 1 \le k \le K \quad (13)$$

Also, the total inventory I_{sk} in tank s during any slot cannot exceed its maximum holding capacity Cap_s^U.

$$I_{sk} \le Cap_s^U \qquad 1 \le s \le S, \, 1 \le k \le K \qquad (14)$$

3.1. Objective

Minimizing Total Costs of FSOP is our scheduling objective. The costs include demurrage costs of VLCCs/ULCCs per hour given by DC_v; costs associated with order delivery delay DCO_o, inventory costs of FSOP-tanks per hour IC_{uk}. Here, d_v represents delay beyond stipulated time of departure of VLCC/ULCC v at the FSOP, d_o represents delivery delay of order o, and $I_{sk=K}$ represents inventory level of tank s during the last slot K.

$$Total\ Costs = \sum_v DC_v.d_v + \sum_{k=K}\sum_s ICs.Isk + \sum_o DCO_o.d_o \tag{15}$$

4. Case Study

We apply our model to 3 VLCC/ULCCs carrying 3, 2 and 2 parcels and arriving at 5, 10 and 15 hours respectively. The FSOP comprises 8 different tanks which deliver 5 possible orders to refineries during the scheduling horizon of 100 hours. We summarize the key aspects of our solution in Fig 1.The solutions were obtained using CPLEX (10) in GAMS (Distribution 22.2) running on a Windows XP PC with AMD Ethlon 64 X2 Dual (3 GHz) processor and with relative optimality gap limit (OPTCR) of 0.0. The model involved 1,965 constraints, 177 binary variables and 7,461 nonzero elements. CPLEX solved the model in 137.95 seconds and gave Total Costs of $10.01.

Fig 1. Key results of Solutions
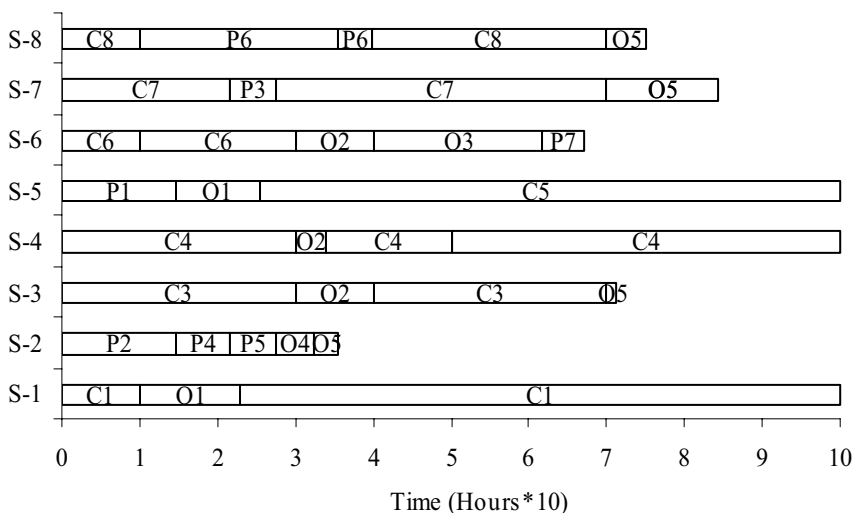
5. Conclusions

This paper introduces a novel concept of crude transshipment using FSOP. To the best of our knowledge, so far no literature exists on the management of such units. The use of asynchronous slot based formulation for crude transshipment is another new aspect it touches. Future work includes making our model more realistic by allowing crude changeovers within tanks.

References

H. Cheng and I.A. Karimi, 2006, Scheduling Transshipment Operations in Maritime Chemical Transportation, Ind. Eng. Chem. Res., 45, 1955-1973.

P.C. Reddy, I.A. Karimi, R. Srinivasan, 2004, A new continuous time formulation for scheduling crude oil operations, Chemical Engineering Science, 59, 1325-1341.

N. Susarla, J. Li, I.A. Karimi, 2008, A Novel Approach to Scheduling Multipurpose Batch Plants Using Unit Slots, Submitted to AIChE Journal.

An Ontological Framework to Support the Implementation of Waste Minimisation Strategies

Adriana P. Reyes-Córdoba[a] , Jorge A. Arizmendi-Sánchez[b]

[a] *School of Chemical Engineering and Analytical Science, The University of Manchester P.O. Box 88, M60 1QD, Manchester, UK.*
[b] *Advantica Ltd. Holywell Park, New Ashby Road, LE11 3BW, Loughborough, England*

Abstract

Modern understanding about waste management has provided a general hierarchy to deal with waste, with a priority in the reduction or minimisation of the amount of waste generated. Different methodologies have been devised to aid the process of implementing waste minimisation in the process industries. However, waste minimisation methodologies are still underestimated or disregarded due to several misconceptions about their applicability. One reason believed to cause this is the lack of an accurate way to identify suitable waste reduction techniques. This arises from a mismatch between the existing methodologies and the information available at the time of the study. An appropriate approach to overcome this situation must consider a more efficient manipulation and analysis of the available information, optimizing the cost and time required to obtain the critical missing data. The structured organisation of the knowledge about a process regarding its description, environmental concerns and available resources provides an effective way to identify the areas which need attention for the purpose of waste minimisation and facilitates the definition of activities that have to be performed to implement the devised solutions.

Keywords: waste minimisation, waste minimisation methodologies, knowledge management, ontological framework

1. Introduction

In process industries, waste refers to any material that cannot be further used in the production of goods. Such material has finished its usable life and has to be disposed properly. Waste minimisation strategies aim at three main goals. The most important of which is to reduce and, when possible, to prevent the generation of waste. However, generation of waste is not always avoidable; thus, the second goal is to improve the quality of any waste generated. This includes reducing the hazard of any waste material so that it can be safely disposed. Additionally, it is important to encourage the reuse, recycling and recovery of any waste material that is generated. The Environment Agency for England and Wales (EA) advises the efficient use of raw materials, energy and water as a way to achieve waste minimisation. In the European Union, it is also considered that processes and activities must be systematically understood and thus changed to prevent and reduce waste.

Figure-1 Main strategy for waste minimisation with associated knowledge requirements methodologies and tools

Several efforts have been made to tackle the complexity of the task that waste minimisation represents. These efforts have been in the form of technical recommendations for operative plants or for the design phase of new plants, however, a challenge is presented in the integration of all the information needed for these approaches to be practical. Figure-1 provides an overview of the general strategy for waste minimisation given by the EA. This overview includes a collection of predominant methodologies studied by different authors [1-4]. These have been integrated into the stage of the EA strategy in which they are most convenient. This figure shows the main concerns about the knowledge required for the effective performance of this strategy. The successful application of these approaches relies on the appropriate identification, gathering and classification of the relevant information required in each case.

Attempts to devise an integrated and optimal strategy to conduct waste minimisation studies usually include the development of automated systems. However, as waste minimisation problems are complex and multidisciplinary, efforts might end up not considering relevant information or overlapping it. This leads to problems like information being presented in incompatible terminology or being structured according to different patterns (i.e. databases, spreadsheets, etc). Knowledge management tools have been recognised as the most practical ways to identify, organize and share this kind of information. Various authors identify the suitability of ontologies for this purpose [5, 6]. Current ontologies in the field of chemical engineering focus on a broad domain and a challenge arises when trying to implement this into waste minimisation efforts. Knowledge is one of the most valuable assets in the effective implementation of waste minimisation programs. Most of the knowledge within a company is known

implicitly to the participating employees. These employees turn into experts on a specific function and become indispensable unless their knowledge is documented. To fully exploit this knowledge it is necessary to make it explicit and shared among the different participants of a specific process [5]. By definition, knowledge management tools are the adequate way to shape knowledge into an explicit form that can be used and shared within an organisation.

2. Knowledge management approach for waste minimisation

2.1. Identification of required domains

The approach presented in this contribution proposes a methodological framework based on ontological principles aimed to support the generation and implementation of waste minimisation strategies. Knowledge is classified in four different domains embedded in an ontological structure: process description, environmental requirements, available documentation and waste minimisation options. This approach enables the classification and retrieval of available information in a more efficient way, providing a more structured framework to support the implementation of waste minimisation strategies. The proposed ontological structure aims to provide answers to these questions: which environmental requirements are not being complied, what kind of information is needed to work towards compliance, how this information is obtained, what kinds of waste minimisation options are available and which knowledge resources are needed and available to implement them. The required domains are constructed by identifying and collecting expert knowledge. Following the general recommendations for ontology construction [7], the process knowledge domain includes a fraction of an existing ontology (OntoCAPE [8]) as the keystone on which further knowledge can be added. The ontological framework proposed bases the process description on the conceptual layer presented in the ontology as the module named *chemical process system*. The main focus is on the actual plant model and the *process*, *process step* and *process units* on which the plant depends. The legislation domain takes into account the conditions set by the IPPC directive for the European Union to promote the use of Best Available Techniques (BAT). These techniques provide the main issues to address for specific sectors within the process industry. The third domain classifies waste minimisation options and methodologies currently in use. These are drawn from the BAT reference documents and from the existing strategies mentioned previously. The knowledge resources domain is constructed by considering the experts required to conduct the waste minimisation strategies which constitute the explicit and tacit forms of knowledge available for an organisation. Additionally, explicit forms of knowledge are taken into account and divided into different tools such as documented information, specialized software, calculations, and external resources.

2.2. Development of the domains

The development of these domains is performed following the same principles of ontological construction. This approach was chosen due to the need for an integrated description of these domains which can be achieved by identifying their key concepts and relationships in the same manner as an ontology. The first stage is to select the scope of the domain and analyze its requirements. This is done by the use of guidance questions that involve queries which the domain should be able to answer. This has the intention of identifying the major requirements, purpose and scope for the domain. The second stage involves the collection of reusable resources such as existing ontological domains and conceptual data models relevant to the scope and to identify parts of

relevance that could be reused The third stage requires building the structure of the domain. This involves identifying important terms that describe the objects and concepts found in the domain, grouping these terms into classes and setting them to a hierarchy with defined relationships. The final purpose is to communicate the conceptualization of the four domains and to provide the backbone for the development of an ontology based on these domains.
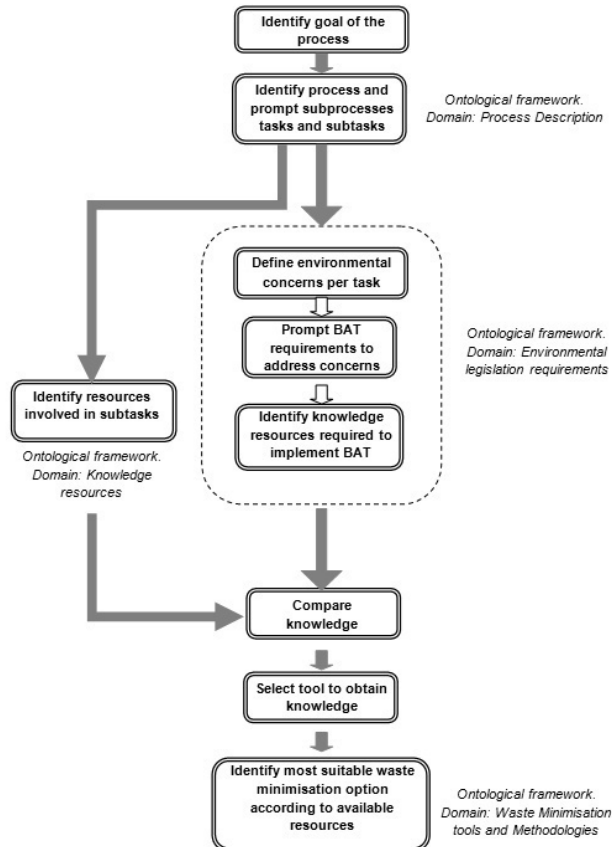
Figure-2 General methodology for the implementation of an ontological framework to aid in waste minimisation

3. Methodology to assess the suitability of waste minimisation actions

The methodology, presented in

Figure-2, starts with the identification of the goal of the process under study. This refers to the crucial knowledge element which dictates what the final purpose of the whole process is. According to this goal it is possible to identify to which industry the process belongs to and which subtasks are involved in it. Once this is done, the environmental requirements for each of the previously identified subprocesses are declared. These are drawn from the environmental legislation domain. Compliance of these environmental requirements is achieved by planning specific actions which use specific information resources. The next step in the methodology is to assess if these information resources

are present within the organisation. To achieve this, the knowledge resources domain is analysed. This domain lists the different information resources that should be available to the company in order to match them with the ones available in reality. In this way the missing knowledge is identified along with the effective sources to find it. Finally, once this knowledge has been gathered it is possible to suggest a waste minimisation option that is practical with the available resources.

4. Use of the framework

The framework has been tested on the case study of the reaction step of phthalic anhydride production. The previously discussed domains provide the framework to study the process as shown in Figure 3.
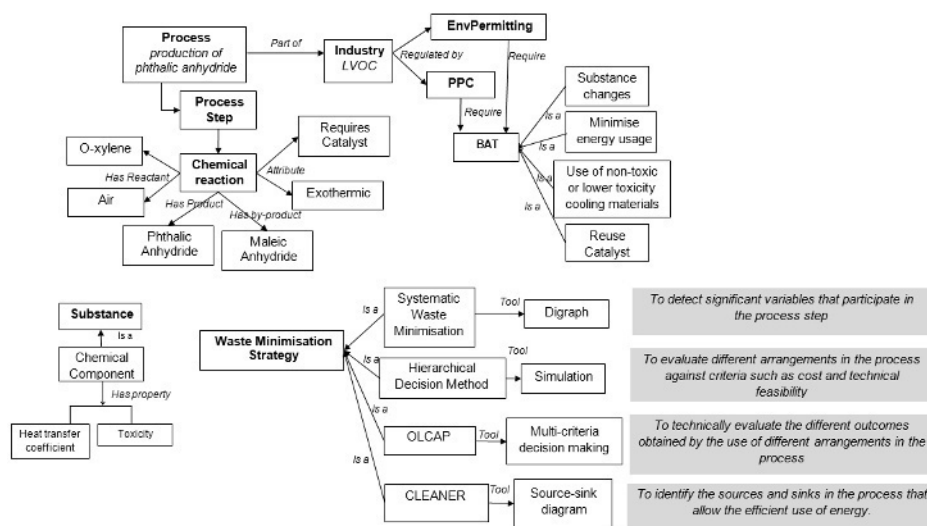
Figure 3 Study of the process using the proposed domains

The generic process steps defined by the process description domain include reaction, separation, and enthalpy change, amongst others. The focus in this case is the reaction process step. The reaction is defined by specific attributes which in this case include the need for a catalyst and the exothermic nature. The reaction also involves the use of specific materials namely o-xylene, air, and a catalyst. The overall process is classified in the large volume organic chemicals (LVOC) industry which is regulated by the Environmental Permitting regulations and the IPPC. Under these regulations the use of best available techniques (BAT) is required. Generic BAT recommendations for the LVOC industry include: minimization of energy usage, regeneration or reuse of catalysts, use of non-toxic or lower toxicity cooling materials amongst others. The exothermic nature of the reaction allows for the first BAT recommendation to be explored. The reaction can be analysed using the identification of process variables and their functionality studied in the waste minimisation strategies that form the waste minimisation domain. The variables identified for the reactor are temperature, pressure and composition. The need to control Temperature is evident and thus the first heuristic rule arises which identifies the action of adding a cooling agent as a solution. The cooling agent has to be evaluated against constraints such as cost, environmental impact

and technical feasibility as determined in the waste minimisation domain. Three different materials are available in this case, mercury, sodium potassium alloy and a heat transfer salt mixture. The material evaluation gives the third material as the best compromise solution between costs, technical feasibility and environmental impact. Furthermore the elaboration of source-sink diagrams (as suggested in the waste minimisation strategies) allows the identification of possible ways of using the heat gained by the salt as energy source for heating process water required for another stage of the process. Knowledge resources required for the evaluation include specialized simulation software, databases with material characteristics (including toxicity, heat transfer coefficient and costs), energy balances and personnel with experience in this kind of studies.

5. Discussion

Organizing the knowledge about a process regarding its description, environmental concerns and available information resources provides an effective way to identify the areas which need attention for the purpose of waste minimisation. This facilitates the process of defining the activities that have to be performed to implement the devised solutions. The assessment and pre-assessment stages of waste minimisation studies are crucial for the successful achievement of the goals set. These stages require vast amounts of information to be available and accurate. They also require multidisciplinary teams with experts and skilled staff from all the areas involved. These resources must be adequately managed in order for them to work towards the final goal of waste minimisation. Effective organisation of available information and knowledge is the key for the proper application of waste minimisation methodologies. Integration of waste minimisation into the field of knowledge management through the use of an ontological framework has been proposed. It has been recognised that ontologies can be used to organize knowledge and therefore tackle different areas without disregarding relevant issues. Waste minimisation methodologies need a tool that considers the complexity of the knowledge that has to be gathered for their correct performance. A systematic approach to organize this knowledge will therefore result in a more practical way to implement current waste minimization methodologies.

References

1. Azapagic, A. and R. Clift, *The application of life cycle assessment to process optimisation.* Computers and Chemical Engineering, 1999. **23**: p. 1509-1526.
2. Halim, I. and R. Srinivasan, *Systematic Waste Minimization in Chemical Processes. 1. Methodology.* Industrial and Engineering Chemical Research, 2002(41): p. 196-207.
3. Mulholland, K. and J. Dyer, *Follow this path to pollution prevention.* Chemical Engineering Progress, 1998. **94**(1): p. 34-42.
4. Dunn, R. and G. Bush, *Using process integration technology for CLEANER production.* Journal of Cleaner Production, 2000. **9**: p. 1-23.
5. Brandt, S.C., et al., *An ontology-based approach to knowledge management in design processes.* Computers and Chemical Engineering, 2008. **32**: p. 320-342.
6. Zhao, C., et al., *Toward intelligent decision support for pharmaceutical product development.* Journal of Pharmaceutical Innovation, 2006: p. 23-35.
7. Gómez-Pérez, A., M. Fernández-López, and O. Corcho, *Ontological Engineering.* 1st ed, ed. X. Wu and L. Jain. 2004: Springer-Verlag London Limited 2004. 403.
8. Morbach, J., A. Yang, and W. Marquardt, *OntoCAPE- A large-scale ontology for chemical process engineering.* Engineering applications of artificial intelligence, 2007. **20**: p. 147-161.

Abnormal Situation Management in a Refinery Supply Chain Supported by an Agent-Based Simulation Model

Koen H. van Dam[a], Zofia Lukszo[a], Rajagopalan Srinivasan[b]

[a] *Department of Energy and Industry, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, the Netherlands*
[b] *Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576, Singapore*

Abstract

Oil refineries are of high importance for global economic health and energy supply; any disruptions to their operations may have major worldwide impact. In this paper the application of an agent-based refinery supply chain model to abnormal situation management is described. Agents represent the various decision makers in the supply chain. They own, operate and manage the elements of the physical network of the supply chain. A disruption in ship arrival is used to illustrate the applicability of the decision support system. The decision support system derives a suitable course of action for a given situation based on the outcomes of a number of simulation runs according to the Nelder-Mead zero-order optimization method. This method is based on identification of the best, the worst, and the second worst outcomes in each iteration for the pre-defined experiment. It can be concluded that the decision support system can interact with multiple actors in the supply chain to diagnose and compensate for unanticipated disruptions, with a substantial impact on refinery productivity.

Keywords: refinery supply chain, agent-based model, decision support, disruption, Nelder-Mead

1. Introduction

Abnormal situations in supply chains (SCs) encompass a range of events outside the "normal" operating modes, including human error, fires, delays in ship arrival, (unplanned) maintenance and equipment failure. As a consequence, planned production targets may not be reached, unless swift response is taken to minimise the negative effects. Which response is the most efficient one, however, is not easy to determine and requires decision support tools. This paper describes how a decision support tool using an agent-based simulation model can help an oil refinery in dealing with disturbances and to ensure smooth operation at minimal costs.

Oil refineries are of high importance for global economic health and energy supply; any disruption to their operations may have major local and worldwide impact. A typical refinery SC comprises crude oil suppliers, 3rd party logistics providers, shippers, jetty operators, the refinery and customers. The refinery occupies a pivotal position in the SC with its functional departments initiating and controlling the interactions with the external entities for the various SC activities, such as procurement, storage, logistics and operations. A maze of complex interactions between the different entities and resulting decisions ensure the orderly and efficient functioning of the supply chain as described in detail by Julka *et al*. (2002) and Pitty *et al*. (2008). The combined

performance determines the economics via crude costs, product prices, operation costs, transportation, etc.

The multi-actor, distributed, complex and dynamic nature of a supply chain can be best evaluated using simulation models. There is a strong need for models that can help decision makers in the process industry to analyse the risks of abnormal situations in the supply chain and to assess possible solutions. The agent-based paradigm, with agents as problem solving entities characterised by autonomy, reactiveness, pro-activeness, goal orientation and the ability to communicate with other agents and the environment (Jennings, 2000), is suitable for supply chain modelling: flexible models can be developed that are easy to extend and which can capture and visualise the complex social interactions between the supply chain units, modelled as agents (van Dam *et al.*, in print). In this paper it is demonstrated how an agent-based model can be used to support decision makers dealing with abnormal situations. While each one of the actors (i.e. stakeholders) can be considered as a main problem owner, here the perspective of the oil refinery is chosen. The decision support tool described here is designed for one such decision maker in an industrial setting.

This paper contributes to the area of the abnormal situation management of a refinery supply chain and to the application of agent-based models to industrial decision making. Section 2 describes a generic decision problem for a refinery supply chain. In section 3 the search algorithm supporting the decision problem is discussed, followed by a specific example in Section 4. Finally, Section 5 gives concluding remarks on the use of agent-based models for decision support.

2. Decision problem

This section describes a decision problem for an oil refinery supply chain. First, the disturbances are defined, followed by the degrees of freedom and the performance criterion.

2.1. Disturbances

Although the approach proposed in this work is general, in this paper the scope is limited to disturbances dealing with the *supply* of crude oil to the crude distillation units. A disruption in the supply can be caused by a delay in the shipment of crudes from the supplier (at a large distance from the refinery) or problems in the tank farm, for example. In both cases the operations department runs the risk of not having access to enough crude to perform the scheduled operations.

These disturbances are then defined as follows:

$\underline{d}$ = *(ShipDelay, StorageProblem)*

ShipDelay $\in$ *{0, 1, 2,..., n}*   in days, for 1 ship for 1 cycle

StorageProblem $\in$ *{0, 1}^{m}*   for each of the *m = 5* crude storage tanks

For instance, $\underline{d}$ = (2, 0,0,0,0,0) indicates that a ship at sea is delayed by 2 days, and there are no problems with the storage installation. Note that *n* for a ship delay can be larger than the time horizon used, effectively "sinking" the ship. For simplicity, we assume that the magnitude of the disturbance is known as soon as the disturbance occurs. In reality, this may involve uncertainty. Furthermore, the granularity could be adjusted so that a delay could be expressed in parts of a day (e.g. hours) in stead of full days.

2.2. Degrees of freedom

When faced with such a disturbance, the problem owner has to make a number of choices. Firstly, he has to determine if the disturbance has a significant effect on the operation of the supply chain. If the effect is deemed minor, no action may be necessary, but when not able to execute the previously planned schedules due to inadequate crude, corrective action may be required. This can be addressed by changing the operating mode, the throughput or by emergency crude procurement. Often a combination of these actions may be needed.

For the Emergency Procurement *EmPr* the procurement department can contact a local supplier to buy crude at a much higher price but with a short lead time. The degree of freedom for each of the five crudes is between 0 kbbl to the amount that could reasonably available on short notice, which is assumed 600 kbbl here. The procurement department has to ask logistics department for the expected delay to be able to make this decision. Note that when there is an error in a storage installation (*StorageInstallationError=1*), emergency procurement will not solve a disturbance, as all crudes have to rest in storage to allow the brine to settle and cannot be transferred directly from the vessel to the crude distillation unit but have to pass storage tanks.

Furthermore, the operations department can choose to Change the Operational Configuration (*COC*), meaning that a different recipe (one out of four) is selected using crudes that are still in stock, but resulting in yields that are not ideal compared to the scheduled operation. Finally, the operations department can Change the Operational Scale (*COS*), to run the refinery at a lower throughput (from 40% to 100% of the maximum capacity, with a minimum throughput to keep plant running), producing less end products but avoiding having to shutdown the plant when crude runs out (or postponing plant shutdown, for example to allow emergency procurement crudes to arrive).

The degrees of freedom are defined as follows:

$$\underline{x} = (EmPr_1, EmPr_2, EmPr_3, EmPr_4, EmPr_5, COC, COS)$$

$0 \leq EmPr_i \leq 600$ 　　　　　　　in kbbl, for each of the $i = 5$ crudes

$COC \; \epsilon \; \{R1, R2, R3, R4\}$ 　　　　discrete choice between operating modes

$40 \leq COS \leq 100$ 　　　　　　percentage of CDU throughput capacity

2.3. Performance criterion

There are different options for the criteria with which to choose the best alternative. As examples, one can look at the overall profit of the refinery (for a certain time frame), profit during the production cycle effected by the disturbance, other financial measures, but also non-economical criteria such as customer satisfaction. Here we choose profit *P* of refinery 14 days after a disruption took place. This means that the effect of a disturbance at *t = 32* will be simulated over the next two cycles of operation during which new raw material deliveries and product dispatch. We expect that the impact of the disturbance would have worn off by then.

The objective function is defined as follows:

$$max\, P(\,\underline{x},\underline{d}\,) = \sum_{t=1}^{50}(\,Income^t{}_{sales}(\,\underline{x},\underline{d}\,) - Cost^t{}_{procurement}(\,\underline{x},\underline{d}\,) - Cost^t{}_{transp}(\,\underline{x},\underline{d}\,) - Cost^t{}_{maint}(\,\underline{x},\underline{d}\,))$$
$$+\, Value_{producstocka}^{t=50}(\,\underline{x},\underline{d}\,) + Value_{rawmaterials}^{t=50}(\,\underline{x},\underline{d}\,)$$

In this equation the value of the product inventories and raw materials at the end of the simulation run have been included. The consequences of the disruption on future cycles are included in the cost function (e.g. if the response is to switch to another mode of operation without any emergency procurement, it is possible that during a later cycle the planned operation cannot be met) but no new decisions following to any such new disturbances are assumed; a single response is formulated.

The transportation costs, and therefore profit, are discontinuous because they are a function of the amount of crude procured, the capacity of the ships (either a VLCC for long distance shipping or a general purpose tanker with much smaller capacity for short haul in the case of emergency procurement) and the travel time. The transport cost is calculated per vessel and the cost for procurement plus transport therefore follows a saw-tooth pattern, making it more difficult to determine the right amount to buy (especially in combination with other measures such as switching to another recipe).

3. Search method for decision support

The decision support system presented here derives the advice for a given situation based on the outcomes of a number of simulation runs according to the Nelder-Mead zero-order optimization method (Nelder & Mead, 1965). The term zero-order refers to the fact that the search of the optimum is carried out without calculating any derivatives of the performance criterion but directly by measuring (without help from a process model) or simulating the state of the system. Zero-order search is recommended when the process has one or more of the following properties (Wright, 1995):

- Process model is difficult or expensive to obtain;
- Process exhibits discontinuities;
- Measurement data are contaminated by significant noise.

The search with a Nelder–Mead simplex algorithm is applicable to solve decision problem formulated in this paper. The method is based on identification of the best, the worst, and the second worst outcomes in each iteration for the pre-defined simplex (a set of experiments). An initial simplex S is defined as a convex hull with $n + 1$ vertices $\{\underline{x}_j\}^{n+1}_{j=1}$ in an n-dimensional space $\mathbb{R}^n$ (in this case $n=7$). These vertices satisfy the non-degeneracy condition, meaning that the volume of the simplex hull is non-zero. For every next iteration $j+1$, the values for $\{\underline{x}_j\}^{n+1}_{j+1}$ are determined by comparing the objective-function values followed by replacement of the worst vertex by another point. The simplex adapts itself to the local landscape and contracts finally on the optimum.

Compared with traditional iterative search algorithms, the proposed examples in this paper show that the pattern search algorithm is robust, easily implemented and easy to communicate to decision makers. Moreover, the search algorithm is more likely to give a global minimum in dealing with nonlinear problems, non-linear constraints and multi-criteria problems (Verwater-Lukszo, 1996; Lu, 2008; Ouria, 2009; Ye 2008).

4. Abnormal situation management supported by an agent-based model

The system model of an oil refinery SC used here for simulation and experimentation to support the decision-making process is an agent-based model developed in Repast, as described in detail in van Dam *et al.*, (2008). Agents model all stakeholders, each with their own unique behavioural rules and own goals. They communicate with each other to make decisions about the trade of (raw) materials or the operation of the refinery, for example. The model has been benchmarked with an equation based model of the same
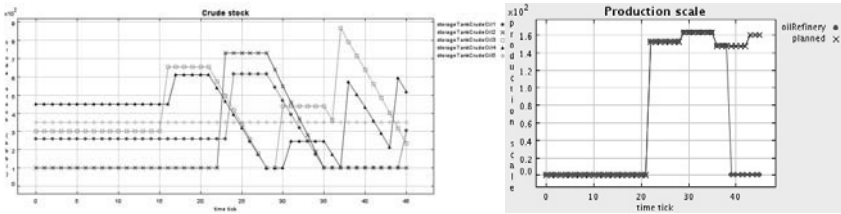
Figure 1. Crude stocks under normal operation (left) and the gap between planned and actual throughput after a disturbance causing a shut-down of the refinery (right)

Table 1. Outcome of the Nelder-Mead simplex optimization algorithm with the agent-based model

	$EmPr_1$	$EmPr_2$	$EmPr_3$	$EmPr_4$	$EmPr_5$	COS	$P(\underline{x},\underline{d})$
1	0	0	0	0	0	60	-5.69 E8
2	0	0	100	100	0	60	-3.65 E8
3	0	0	200	300	0	60	-1.61 E8
4	0	0	300	250	0	55	-1.64 E8
5	0	0	300	250	0	50	-6.19 E7
6	100	100	100	100	100	50	-3.70 E8
7	10	10	500	600	10	50	1.39 E8
Final	4	4	349	382	4	54	1.45 E8

system through replication (van Dam *et al.*, in print). It has been used for the assessment of various policies, including those for procurement and storage.

To determine *nominal* process conditions the model performs optimisation for normal operation by choosing which crudes to buy, how much crude is needed and from which supplier to order. The mode of operation is scheduled based on predicted demands and the throughput for operation of the refinery is set based on actual demand from the consumer. When a disturbance is manifesting itself this normal approach is not adequate any more because the complexity is too big and there are too many inter-depended variables. A model-based decision-support tool is therefore called for.

Figure 1 illustrates the refinery behaviour under normal operation and after disturbance $\underline{d}$ = (30,0,0,0,0,0) which occurs on day t = 22, resulting in a loss of 24 m$ because of loss of production. Next, the agent-based simulation model supports the choice which response is the most appropriate given the many degrees of freedom. Preliminary results for the decision on emergency procurement and the change of operational scale are shown in Table 1. After 20 iterations no further improvement is made so the algorithm is stopped. The proposed solution prevents a shut-down of the refinery by buying emergency crudes to make up for the delayed ship and by slightly reducing the throughput. The loss caused by the disruption is reduced by 14.7 m$ (not keeping in mind penalties to be paid by the shipper for delays).

The change in operational configuration was not included here. For each possible *COC ϵ {R1, R2, R3, R4}* an optimisation for $EmPr_i$ and *COS* has to be performed because choosing a different recipe will influence the criterion surface. CPU time is not an obstacle to running the optimisation through multiple cycles, after which the best solution can finally be selected.

5. Final remarks

The Nelder-Mead optimisation method is a commonly used method in process systems engineering using mathematical models or samples from experiments in a real system.

Here we demonstrated that it can be a powerful approach when combined with an agent-based model, too. The model is developed in a bottom-up fashion, making it relatively simple to change the configuration: It is easy to include new actors (e.g. more suppliers with different lead times) or to adjust the physical configuration (e.g. of the storage tanks). ABM are particularly suitable to experiment with different scenarios and to answer "what if" questions, critical for decision support under disruptions.

Determining the right combination of options (e.g. a switch of operational configuration and emergency procurement for the crudes used for that recipe instead of the ones in the delayed tanker) is difficult, and becomes extremely hard when responses at different times are allowed (e.g. a small emergency procurement now, switching recipe a few days later, switching back when the delayed ship arrives, etc) and when responses to new disturbances are also included (e.g. the long term effects of not always following planned operation). The full power of the approach presented here becomes visible then.

An illustrative case study using an agent-based simulation model of a supply chain with the Nelder-Mead optimisation method was presented and it was demonstrated how agent-based models can be applied as decision support tool. The approach presented can provide valuable support choosing the right response to abnormal situations in such a highly complex system.

Acknowledgment

References

K. H. van Dam, A. Adhitya, R. Srinivasan, Z. Lukszo, 2008, Benchmarking numerical and agent-based models of an oil refinery supply chain, ESCAPE-18, France.

K. H. van Dam, A. Adhitya, R. Srinivasan, Z. Lukszo, in print, Critical evaluation of paradigms for modelling integrated supply chains, accepted for J. Computers and Chemical Engineering

N. Jennings (2000). On agent based software engineering, Artificial Intelligence 117: 277-296.

N. Julka, I. Karimi, R. Srinivasan, 2002, Agent-based supply chain management – 2: a refinery application, J. Computers and Chemical Engineering, 26(12), pp. 1771–1781.

L. Lu, D.J. Murray-Smith, D.G. Thomson, 2008, Issues of numerical accuracy and stability in inverse simulation, Simulation Modelling Practice and Theory 16, pp.1350–1364.

J.A. Nelder, R. Mead, 1965, A Simplex Method for Function Minimization, Comput. J. 7, pp.308–313.

A. Ouria, M. M. Toufigh, 2009, Application of Nelder-Mead simplex method for unconfined seepage problems, Applied Mathematical Modelling (In Press).

S. Pitty, W. Li, A. Adhitya, R. Srinivasan, and I.A. Karimi, 2008, Decision Support for Integrated Refinery Supply Chains. 1. Dynamic Simulation, J. Computers and Chemical Engineering, Vol 32, pp 2767–2786.

Z. Verwater-Lukszo, 1996, A practical approach to recipe improvement and optimization in the batch processing industry, PhD Thesis, Eindhoven University of Technology

M. H. Wright, 1995, Direct search methods: Once scorned, now respectable, In: D. F. Griffiths, & G. A. Watson (Eds.), Numerical analysis, Harlow, UK: Addison-Wesley, Longman.

Q. Xiong, A. Jutan, 2003, Continuous optimization using a dynamic simplex method, Chemical Engineering Science 58, pp. 3817–3828.

T. Ye, C.H. Xiong, 2008, Geometric parameter optimization in multi-axis machining, Computer-Aided Design 40, pp. 879–890.

Detailed Supply Chain Design Considering Production Campaigns

Yanina Fumero,[a] Gabriela Corsano,[a] Jorge M. Montagna[a,b]

[a]*INGAR, Avellaneda 3657, (3000) Santa Fe, Argentina*
[b]*CIDISI – Facultad Regional Santa Fe, UTN, Santa Fe, Argentina*

Abstract

Until now, supply chain (SC) design models have been mainly focused on SC integration, where nodes allocation and links among them are selected in order to allow an efficient operation of the whole system. Usually, the detailed configuration and operation of the plants have not been taken into account. In this work, a heuristic strategy is presented in order to design the SC, including the structure and the operation of plants with mixed product campaigns. The incorporation of plant design and mixed product campaign in the SC design model leads to a non linear formulation. Hence, a two stages approach is addressed in order to solve this problem through linear models. In the first stage, a SC design standard model is solved in order to obtain the network design with minimum logistic cost and the production of each selected plant. These results are used to estimate the possible campaigns composition for each plant. In the second stage, specific scheduling constraints are incorporated in the plant design model to determine the optimal mixed product campaign configuration and the structure of each plant, minimizing the investment cost. This methodology allows obtaining the SC logistic configuration and, for each selected plant, the optimal mixed production campaign simultaneously with the plant design, in order to meet a specified economic criterion fulfilling demand requirements.

Keywords: supply chain, multiproduct batch plants design, mixed product campaign, optimization.

1. Introduction

A Supply Chain (SC) is a network of organizations that performs a set of activities in different phases of the production process. Determining an optimal configuration and efficient operation of the SC is a complex problem, taking into account the different criteria involved (Amaro and Barbosa-Póvoa, 2008; Guillén et al. 2006). Following Shah (2005), there are no published works dealing with the connection between plant and SC design and operation. Usually, plants are modeled as black boxes and nodes allocation and flows among them are determined in the solution.

In a previous work (Corsano and Montagna, 2008), a new approach has been presented in order to include plant design in SC models for multiproduct batch plants. This work shows that SC costs, design and capacity are severely affected by the plant structure. However, that work assumes single product campaigns (SPC) in order to simplify the formulation. This assumption is usual in these models but it is very limiting from the commercial point of view where a more steady supply of products is required.

At plant level, Birewar and Grossmann (1989) consider that the sequencing of batches of different products in mixed product campaign (MPC) can reduce idle times to increase the utilization of equipment. Nevertheless, their models incorporate MPC in the

design of multiproduct batch plants considering only one unit per processing stage. On the other hand, parallel units can reduce the cycle time of the campaign. However, there are few works considering MPC taking into account the formulation is more complex.

In this work, a heuristic approach is presented for SC and plants design optimization considering MPC. The proposed methodology is structured in two stages: first a MILP model determines the SC design and then, according to the production amounts of each product in each plant, a MILP model is used to select the MPC and the structure of each plant with minimum investment cost. The model assumes that units are provided using discrete sizes following the commercial procurement policy and that the transfer policy between stages is zero-wait (ZW). Unit duplication out of phase for each stage is taken into account in order to achieve an efficient plant design.

2. Problem Description and Mathematical Model

In this work, the SC comprises four echelons: raw materials sites, plants, warehouses and customers zones. At each raw material site, several types of raw materials are available to be delivered to plants. Each product can be produced at several plants. The warehouses can be supplied from more than one plant. Each customer zone demands for one or more products. The problem is to determine the allocation of raw materials sites, plants and warehouses, the flows among them, the amounts to be produced in each plant, and the design and the optimal MPC of each selected plant in order to meet a specified economic criterion fulfilling demand requirements during a time horizon.

A decomposition strategy is addressed through two MILP models where, in the first stage, a general model is formulated for the SC design to determine: i) the allocation of raw material sites, plants and warehouses; ii) the production of each product in each plant; and iii) the flows among SC nodes; in order to minimize the logistic cost (transport cost from raw materials sites to plants, from plants to warehouses and from warehouses to customer zones) and inventory costs. From the optimal solution, relations among the number of batches of each product for selected plants can be established according to the obtained production amounts. Therefore, the possible campaign compositions can be estimated.

Then, for each selected plant and each proposed composition, a MILP model is formulated. The production sequence of the MPC and the plant structure that minimize the investment cost are attained. Several compositions of the production campaign can be proposed for each plant, and the MPC with smallest objective function value is selected as the optimal one.

2.1. Logistic Model

Given the product demands, the set of available raw materials and their availability at each raw material sites, and the transport and inventory costs, the MILP model determines the production of each product in each plant, the allocation of plants, raw material sites and warehouses, and the flows among them in order to achieve the most economical design. Basically, the model considers mass balances between the SC nodes and allocation constraints for plants and warehouses. Similar constraints are presented by Corsano and Montagna (2008).

2.2. Plant Design and Mixed Product Campaign Model

This model introduces the MPC sequencing in the batch plant design. Previous models assumed SPC that simplified the formulation. From the logistic model solution, for each installed plant, the relation among the number of batches of the different products can be estimated. Therefore, the campaign composition can be proposed for each plant specifying the number of batches of each product in the campaign as data. Batches are

sequenced using the concept of production slot (Birewar and Grossmann, 1989), which requires the assignment of batches to these slots.

A batch plant includes N stages, K_j available units in parallel in stage j to produce N_p products. N_b represents the number of batches of the proposed MPC for that plant, NBC_i is the number of batches of product i in the campaign and N_s is the number of production slots where the N_b batches will be assigned. Since the parallel units in each stage j are assumed to be identical, a batch can be processed on any unit and the processing time will be the same. Thus, a unique processing time t_{ij} is defined for product i in the stage j.

2.2.1. Allocation Constraints

Binary variables Y_{bi} are used to indicate that a batch b is of product i. The following constraints must be satisfied for these variables:

Each batch b corresponds exactly to one product i: $\quad \sum_i Y_{bi} = 1, \qquad \forall b \qquad$ (1)

Product i is assigned to NBC_i batches: $\quad \sum_b Y_{bi} = NBC_i, \qquad \forall i \qquad$ (2)

In order to determine the processing of batch b in stage j, the binary variable YY_{bjkl} is defined and it takes value 1 if batch b is assigned to slot l and is processed in unit k in stage j, and 0 otherwise.

The slot l must be processed in a unit of stage j: $\quad \sum_b \sum_k YY_{bjkl} = 1, \qquad \forall j, l \qquad$ (3)

Moreover, in order to eliminate alternative solutions, for each stage j, the slot l_1 is arranged in unit k_1: $\qquad \sum_b YY_{bjkl_1} = 0, \qquad \forall j, k, \; k \geq k_2 \qquad$ (4)

Each batch b is processed in only one slot l of only one unit k of stage j:

$$\sum_l \sum_k YY_{bjkl} = 1, \qquad \forall b, j \qquad (5)$$

In addition, each batch b on different stages must be assigned to the same slot l:

$$\sum_l \sum_k l\, YY_{bjkl} = \sum_l \sum_k l\, YY_{bj'kl} \qquad \forall b, j, j', \; j < j' \qquad (6)$$

Let Z_{jk} be the binary variable equal to 1 if unit k of stage j is employed. In order to reduce the search space, units will be utilized in ascending order, i.e. unit $k+1$ is used only if unit k has been already allocated: $\qquad Z_{jk} \geq Z_{jk+1}, \qquad \forall j, k \qquad$ (7)

When unit k of stage j is utilized, then at least a batch in a slot must be processed in that unit. Otherwise, this unit is not utilized to process any batch:

$$\sum_l \sum_b YY_{bjkl} \geq Z_{jk}, \qquad \forall j, k \qquad (8)$$

$$YY_{bjkl} \leq Z_{jk}, \qquad \forall j, k, l, b \qquad (9)$$

2.2.2. Timing constraints

Let T_{bj} be the processing time for batch b in stage j, TI_{jkl} and TF_{jkl} the starting and finishing times, respectively, to process the slot l in unit k of stage j. Equation (10) establishes the relation between these variables and YY_{bjkl}:

$$TF_{jkl} = TI_{jkl} + \sum_b \sum_i t_{ij} Y_{bi} YY_{bjkl}, \qquad \forall j, k, l \qquad (10)$$

When no batch is assigned to slot l of unit k in stage j, $YY_{bjkl} = 0$, and TI_{jkl} and TF_{jkl} are equal. This constraint can be linearized introducing an extra nonnegative variable W_{bijkl}

that takes value 1 if $Y_{bi} = 1$ and $YY_{bjkl} = 1$, and 0 otherwise, so (10) is replaced by (11):

$$TF_{jkl} = TI_{jkl} + \sum_b \sum_i t_{ij} W_{bijkl}, \qquad \forall j, k, l \tag{11}$$

where $TI_{j,k,l_1} = 0$ and $W_{bijkl} \geq Y_{bi} + YY_{bjkl} - 1, \qquad \forall b, i, j, k, l$. $\tag{12}$

The new variable must satisfy the following constraint:

$$\sum_k \sum_b \sum_l W_{bijkl} = NBC_i, \qquad \forall i, j \tag{13}$$

In order to avoid the overlapping between different batches in a unit, the following constraint is added: $TF_{jkl} \leq TI_{jkl'}, \qquad \forall j, k, l, l', \; l < l' \tag{14}$

If no batch is processed in slot $l+1$, then the starting time of this slot is equal to final time of slot l: $TF_{jkl} - TI_{jkl+1} \geq -M \sum_b YY_{bjkl+1}, \qquad \forall j, k, l \tag{15}$

where scalar M is large enough.

When the batch b utilizes units k in stage j and k' in stage $j+1$, the ZW policy is imposed by: $TF_{jkl} = TI_{j+1k'l}, \qquad \forall j, l, k, k' \tag{16}$

In order to obtain the cycle time of the campaign, CT, the cycle time of unit k of stage j, CT_{jk}, must be determined, where l_1 and l_n are the first and the last slots of stage j:

$$CT_{jk} = TF_{jkl_n} - TI_{jkl_1}, \qquad \forall j, k = k_1 \tag{17}$$

$$CT_{jk} = TF_{jkl_n} - TI_{jk\tilde{l}}, \quad \forall j, k, k \geq k_2 \quad \text{where} \; \tilde{l} = \min_l \left\{ \sum_b YY_{bjkl} = 1 \right\}. \tag{18}$$

Thus, CT is given by: $CT \geq CT_{jk}, \quad \forall j, k$. $\tag{19}$

2.2.3. Design constraint

Let $SV_j = \{VF_{j1}, \ldots, VF_{jP}\}$ be the set of available discrete sizes for stage j. vv_{jp} is a binary variable equal to 1 if units of stage j have size p, with $p=1,\ldots,P$. Then, the size of units of stage j, V_j, is given by: $V_j = \sum_p vv_{jp} VF_{jp}, \qquad \forall j \tag{20}$

where $\sum_p vv_{jp} = 1, \qquad \forall j$. $\tag{21}$

SF_{ij} is the size factor of product i in stage j, Q_i the demand of product i, α_j, β_j cost coefficients of units in stage j, and H the time horizon. Let NN be a nonnegative variable that indicates the number of times that the campaign will be repeated in H. The following equations must be satisfied:

$$NBC_i \, NN \, B_i = Q_i, \qquad \forall i \tag{22}$$
$$V_j \geq SF_{ij} B_i, \qquad \forall i, j \tag{23}$$

Substituting equations (20) and (22) into inequality (23), yields:

$$NN \geq \sum_p \frac{SF_{ij} Q_i}{NBC_i VF_{jp}} vv_{jp}, \qquad \forall i, j \tag{24}$$

Defining the time horizon constraint by: $CT \, NN \leq H \tag{25}$
and replacing with (24), the following inequality must be satisfied:

$$\sum_p \frac{SF_{ij} Q_i}{NBC_i VF_{jp}} vv_{jp} CT \leq H, \qquad \forall i, j \tag{26}$$

In order to avoid non linear constraints, a new variable is defined as $X_{jp} = vv_{jp} CT$, and replaced in (26). The following constraints must be also satisfied:

$$\sum_p X_{jp} = CT, \qquad \forall j \tag{27}$$

$$X_{jp} \leq CT^{UP} vv_{jp}, \qquad \forall j,p \tag{28}$$

2.2.4. Objective Function

The objective function minimizes the investment annual cost given by:

$$CI = \sum_j \sum_k Z_{jk} \alpha_j V_j^{\beta_j} \tag{29}$$

Considering equations (20) and (21), and defining the variable XX_{jkp}, which take value 1 if $Z_{jk} = 1$ and $vv_{jp} = 1$, and 0 otherwise, the equation (29) becomes linear:

$$CI = \sum_j \sum_k \sum_p \alpha_j VF_{jp}^{\beta_j} XX_{jkp} \tag{30}$$

This objective function subject to the previous constraints defines the MILP model to obtain simultaneously the optimal campaign sequencing and the plant design.

3. Example

Consider a network with 3 sites which supply 3 raw materials, 3 possible plants with 4 stages which can produce 3 products, 3 possible warehouses and 3 customer zones. The available sizes for units are $SV = \{100, 250, 400, 650, 900\}$ and their cost coefficients are: α_{J1}=5000, α_{J2}=8000, α_{J3}=7000, α_{J4}=6000, $\beta_{J1}=\beta_{J2}=\beta_{J4}$=0.6 and β_{J3}=0.7. Table 1 shows demands in the time horizon H = 7000 hrs, and, Table 2 relevant data for plants.

The optimal solution for the logistic model is illustrated in Fig. 1. Two plants are installed: L1 that produces all products and L3 only two products.

For each plant, the relation among the number of batches of the different products can be estimated. For instance, for the plant L1 the following campaign compositions can be proposed: 2 batches of I1, 1 of I2 and 2 of I3; or 2 of I1, 1 of I2 and 1 of I3; or 4 of I1, 2 of I2 and 3 of I3; among others. The optimal solutions of plant design and MPC sequencing model for the first two compositions are shown in Tables 3 and 4. Up to two parallel units are allowed for every batch stages.

The annual investment cost is \$416400 and the optimal sequence is I3-I2-I3-I1-I1 for the first proposed campaign composition; and \$518000 and I3-I2-I1-I1 for the second one. From the proposed compositions and the reported results, the first MPC campaign is selected as the optimal one, and Fig. 2 shows its Gantt chart. The cycle time of the campaign is equal to 33.33 hrs and the campaign is repeated 210 times.

Table 1.					Table 2. Processing times and Size factors for each plant								
Product demands (x10³kg)						t_{ij} (h)					SF_{ij}		
Product	C1	C2	C3			J1	J2	J3	J4	J1	J2	J3	J4
I1	140	90	100		I1	6	12	4	10	0.7	0.6	0.5	0.7
I2	70	45	50		I2	4	9	6	20	0.6	0.55	0.65	0.6
I3	60	80	60		I3	4	10	3	5	0.7	0.6	0.5	0.6
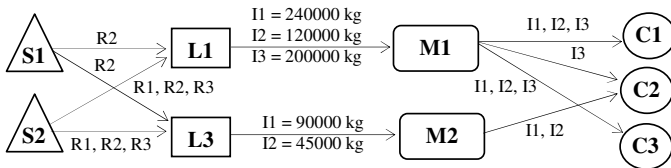
Figure 1. SC optimal design for logistic model

For plant L3, the composition campaign with 2 batches of I1 and 1 of I2 was proposed. For the optimal solution, the campaign configuration is I1-I2-I1 and the plant design consists of only one unit per stage of 250 l. The cycle time of the campaign is equal to 23 hrs and the campaign is repeated 126 times.

The models were solved using GAMS via CPLEX solver, with a 0% optimality gap.

Table 3. Optimal solution for L1 with campaign composition 2 I1 – 1 I2 – 2 I3

Stages	J1	J2	J3	J4
Units Number	1	2	1	2
Units Sizes (l)	400	400	400	400

$B_{I1} = 571.43$　$B_{I2} = 571.43$　$B_{I3} = 476.19$
Optimal Schedule: I3 - I2 - I3 - I1 - I1
Constraints: 695
Binary variables: 228
Continuous variables: 379
CPU time (s): 47.578

Table 4. Optimal solution for L1 with campaign composition 2 I1 – 1 I2 – 1 I3

Stages	J1	J2	J3	J4
Units Number	1	2	1	2
Units Sizes (l)	650	650	400	650

$B_{I1} = 480$　$B_{I2} = 480$　$B_{I3} = 800$
Optimal Schedule: I3 - I2 - I1 - I1
Constraints: 529
Binary variables: 156
Continuous variables: 291
CPU time (s): 8.734
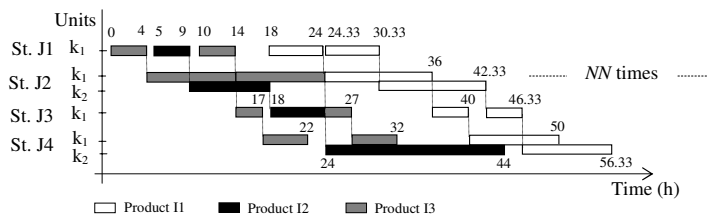
Figure 2. Gantt chart for campaign I3-I2-I3-I1-I1

4. Conclusions

A decomposition strategy for SC design optimization taking into account plants configuration and MPC was presented. In first stage a MILP model determines the SC design and then, according to the production amounts of each product in each plant, a new MILP model is used to determine the structure of each plant with minimum investment cost taking into account the MPC sequencing. Usually MPC sequencing is not considered in multiproduct batch plant design models.

As future work, the authors are exploring a methodology to solve both models in an overall MILP formulation, handling all the requirements.

Due to space inconvenient, some problem data was not reported and they are available for everyone who requests them.

References

A.C.S. Amaro and A.P.F.D. Barbosa-Póvoa, 2008, Supply chin management with optimal scheduling, Ind. Eng. Chem. Res., 47, 116-132.

D. Birewar and I. Grossmann, 1989, Incorporating scheduling in the optimal design of multiproduct batch plants, Comp. & Chem. Engn., 13, 141 – 161.

G. Corsano, and J. M. Montagna, 2008, Mathematical modeling for simultaneous design of plant and supply chain in batch process industry, Under review in Ind. Chem. Eng. Res.

G. Guillén et al., 2006, Addressing the design of chemical supply chain under demand uncertainty. Ind. Eng. Chem. Res., 45, 7566 – 7581.

N. Shah, 2005, Process industry supply chains: Advances and challenges. Comp. & Chem. Engn., 29, 1225 – 1235.

Multi-Site Scheduling/Batching and Production Planning for Batch Process Industries

Georgios M. Kopanos, Luis Puigjaner

Universitat Politècnica de Catalunya-ETSEIB, Diagonal 647, 08028, Barcelona, Spain

Abstract

In the current work, a general mixed-integer linear programming (MILP) model is presented that contemplates most of the key aspects found in the scheduling/batching and production planning for multi-site batch process industries. Batch sizes and processing times are not fixed; instead they are optimized simultaneously with scheduling decisions that also consider products' changeovers and multiple production plants. This way a short-term enterprise-wide optimization entity is built up. Backlogs and period crossover costs/times are also explicitly considered. An illustrative case study is presented to highlight the characteristics of the proposed model.

Keywords: multi-site, scheduling, lot-sizing, planning, MILP.

1. Introduction

Nowadays, manufacturing facilities participating in the same supply chain should intensively share the necessary information and coordinate for the overall planning and scheduling tasks to achieve a holistic optimization performance (Alvarez, 2007). The multi-site production scheduling problem seeks the optimum use and allocation of available resources among several production sites. Scheduling and lot-sizing (batching) are a vital part of operational production planning and control. Moreover, simultaneous lot-sizing and scheduling is essential if sequence-dependent setup costs/times occur during the production process. Several attempts to cope with the scheduling/planning problem can be found in the literature. These contributions are mainly based on discrete or slot-based time representations. By nature, these approaches may provide myopic optimal solutions once implemented; especially, when processing times are lower than the predefined time intervals, since they usually restrict the allocation to only one product per unit and time interval. Thus, the production capacity is underestimated. One way to mitigate this effect, it is to define shorter time intervals or slots, at the expense of increased computational effort that may cause the problem to become intractable. Furthermore, changeovers are not properly handled and high unit idle times may appear. Note that the production scheduling problem has received little attention (Floudas & Lin, 2004). In a multi-site production environment, planning each plant separately will not, in general, lead to optimal solutions. Thus, a multi-site optimization over all production facilities is required (Bassett et. al, 1998; Shah, 1998).

2. Problem statement

The multi-site scheduling/batching and planning problem in batch plants with different processing units in parallel is addressed. Batch-product to unit-plant assignment and batch sequencing meeting a production goal constitute the problem under study. The main problem characteristics and proposed model assumptions include:

- A unit cannot process more than one batch at a time.
- Batch sizes B_{pijn} and processing times pt_{pijn} are variables.
- The unsatisfied demand of a period is not lost. It can be met in next periods generating a backlog cost as long as it is not satisfied.
- Each production facility has a maximum storage capacity.
- Raw materials' flows among plants are allowed PP_p (Fig.1).
- Products' demands, dem_{in}, are realized in specific time intervals, which constitute the production time intervals n.
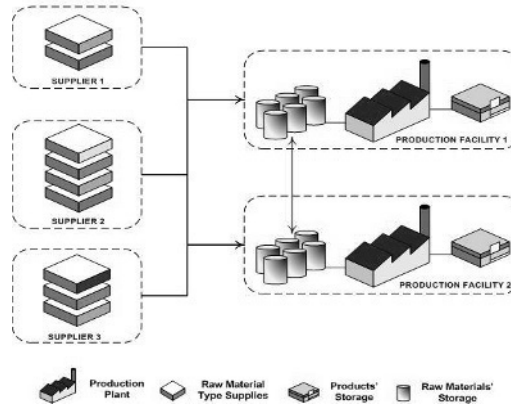
Figure 1. Multi-site production scheme

- Different units j may be installed to each plant p, PJ_j; plants may not be identical.
- Supply of raw materials availability is considered, SR_r; each supplier may provide a different set of raw materials, and with different maximum capacity.
- Final products recipe may need more than one raw material type, IR_r.
- Total sequence-dependent setup times $sd_{ii'j}$ are less than twice the time horizon n.

3. Mathematical model

In the proposed mathematical formulation, the problem constraints have been grouped according to the type of decision (such as lot-sizing, sequencing) to which they are associated. The basic set of equations are stated and briefly explained next.

3.1. Lot-sizing constraints

Parameters ρ_{ij}^{min} and ρ_{ij}^{max} denote the minimum and maximum production rate for producing product i at unit type j, respectively. Set JI_i refers to units of type j that can process product i, while JP_p corresponds to the set of plants p that have installed unit j.

$$\rho_{ij}^{min} pt_{pijn} \leq B_{pijn} \leq \rho_{ij}^{max} pt_{pijn} \qquad \forall p, i, j \in (JP_p \cap JI_i), n \qquad (1)$$

$$pt_{ij}^{min} Y_{pijn} \leq pt_{pijn} \leq pt_{ij}^{max} Y_{pijn} \qquad \forall p, i, j \in (JP_p \cap JI_i), n \qquad (2)$$

Y_{pijn} corresponds to the assignment binary variable of a batch of product i to unit j of plant p while pt_{pijn} and B_{pijn} variables denote the batch size and the processing time of every product-batch i at unit j of plant p at planning period n, respectively.

3.2. Sequencing-allocation constraints

Eqs.(3)-(5) apply the unit-specific immediate precedence sequential notion for the sequencing and allocation decisions. $X_{pii'jn}$ constitutes the sequencing binary variable, which is active, i.e. $X_{pii'jn}=1$, whenever a product-batch i' directly follows a product-batch i in the same unit j at production plant p at planning period n. Eq.(3) denotes that the number of the total number of active sequencing variables in a specific processing unit j is equal to the total number of the assigned products in this unit j minus 1.

$$\sum_{i \in IJ_j} \sum_{i' \neq i, i' \in IJ_j} X_{pi'ijn} = \sum_{i \in IJ_j} Y_{pijn} - 1 \qquad \forall p, j \in JP_p, n \qquad (3)$$

$$\sum_{i'\neq i, i'\in IJ_j} X_{pi'ijn} + WF_{pijn} = Y_{pijn} \qquad \forall p, i, j \in (JP_p \cap JI_i), n \tag{4}$$

$$\sum_{i'\neq i, i'\in IJ_j} X_{pii'jn} + WL_{pijn} = Y_{pijn} \qquad \forall p, i, j \in (JP_p \cap JI_i), n \tag{5}$$

To continue with, WF_{pijn} and WL_{pijn} binary variables represent the first and the last assigned product-batch i at each unit j of every plant p at each planning period n. Set IJ_j contains all final products i that can be processed in unit type j. Eq.(4) states that when a batch i is allocated to a unit j, it may be assigned in the first place, i.e. $WF_{pijn}=1$, or there exists a batch i' that is processed before it, i.e. $X_{pi'ijn}=1$. Following the same logic, eq.(5) expresses that when a batch i is allocated to a unit j, it may be assigned in the last place, i.e. $WL_{pijn}=1$, otherwise it exists a batch i' that is processed after it, i.e. $X_{pii'jn}=1$.

3.3. Timing constraints

It is worth mentioning that eqs.(3)-(5) are not sufficient to prevent the generation of sequence sub-cycles. Hence, the timing constraints of eqs.(6)-(7) are included into the formulation in order to thwart the occurrence of any possible sequence sub-cycle.

$$C_{pi'jn} \geq C_{pijn} + pt_{pi'jn} + su_{i'j}WF_{pi'jn} + sd_{ii'j} - M(1 - X_{pii'jn})$$
$$\forall p, i, i'\neq i, j \in (JP_p \cap JI_i \cap JI_{i'}), n \tag{6}$$

$$su_{ij}WF_{pijn} + pt_{pijn} + \sum_{i'\neq i, i'\in IJ_j} sd_{i'ij}X_{pi'ijn} \leq C_{pijn} \leq Hor_{pjn}Y_{pijn}$$
$$\forall p, i, j \in (JP_p \cap JI_i), n \tag{7}$$

Eq.(6) forces the timing between a pair of two different product-batches i and i' that can be allocated to the same unit j. C_{pijn} variable denotes the completion time of a product-batch i at unit j, and su_{ij} is the sequence-dependent setup time. Finally, eq.(7) provides the upper and the lower bound for the completion time variable C_{pijn}.

3.4. Period crossovers timing constraints

In order to take into account the sequence-dependent costs and times between two consecutive batches across the borders of two adjacent production periods, the following set of equations is needed. By using eq.(8), sequence-dependent ($ssd_{ii'j}$) and –independent ($su_{i'j}$) times are treated in tandem (see Fig.2).

$$sd_{ii'j} = ssd_{ii'j} + su_{i'j} \qquad \forall i, i'\neq i, j \in (JI_i \cap JI_{i'}) \tag{8}$$
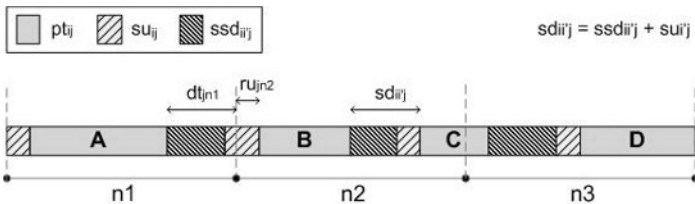
Figure 2. Illustrative scheme for variables dt_{pjn} and ru_{pjn}, and parameter $sd_{ii'j}$.

The product-batch i' that is assigned first at unit j of plant p at period n ($WF_{pi'jn}=1$) follows the last processed product-batch i at unit j of plant p at previous period $n-1$ ($WL_{pijn-1}=1$), as eq.(9) states. Binary variable $X_{pii'jn}^{cros.}$ corresponds to the sequencing variable across the border of two successive planning periods; period $n-1$ and period n.

$$X^{cros.}_{pii'jn} \geq WL_{pijn-1} + WF_{pi'jn} - 1 \quad \forall p, i, i' \neq i, j \in (JP_p \cap JI_i \cap JI_{i'}), n > 1 \quad (9)$$

Additional constraints, eqs.(10)-(11) indicate that the total operating time plus the total changeovers time plus the 'ready' unit time, ru_{pjn}, and the 'dead' unit time, dt_{pjn}, is less than the available unit and plant dependent scheduling horizon Hor_{pjn}. Fig.2 delineates a representative scheme of the continuous variables ru_{pjn} and dt_{pjn}. These variables have been introduced in order to take into account effectively the sequence-(in)dependent setup times between two consecutive batches across periods' borders. By eq.(12), the sequence-(in)dependent setup times can be split in two adjacent planning periods, thus improving capacity utilization.

$$ru_{pjn} + dt_{pjn} + \sum_{i \in IJ_j} pt_{pijn} + \sum_{i \in IJ_j} \sum_{i' \neq i, i' \in IJ_j} sd_{ii'j} X_{pii'jn} \leq Hor_{pjn}$$

$$\forall p, j \in JP_p, n > 1 \quad (10)$$

$$ru^0_{pjn} + dt_{pjn} + \sum_{i \in IJ_j} (pt_{pijn} + su_{ij} WF_{pijn})$$

$$+ \sum_{i \in IJ_j} \sum_{i' \neq i, i' \in IJ_j} sd_{ii'j} X_{pii'jn} \leq Hor_{pjn} \quad \forall p, j \in JP_p, n = 1 \quad (11)$$

$$ru_{pjn} + dt_{pjn-1} \geq \sum_{i \in IJ_j} \sum_{i' \neq i, i' \in IJ_j} sd_{ii'j} X^{cros.}_{pii'jn} \quad \forall p, j \in JP_p, n > 1 \quad (12)$$

The aforementioned set of equations can be viewed as the extended formulation of a recent contribution (Sung & Maravelias, 2008) where only sequence-independent setups were considered.

3.5. Mass balance constraints

Eq.(13) reflects the total production of a product i at each period n, Pr_{in}.

$$Pr_{in} = \sum_p \sum_{j \in (JP_p \cap JI_i)} B_{pijn} \quad \forall i, n \quad (13)$$

Eq.(14) corresponds to the products mass balance at each period n. The unsatisfied demand of a period n, dem_{in}, is represented as backlog, $Back_{in}$. This backlog is added to the demand of the following period; it also generates an additional cost. St_{pin} stands for the stocked amount of product i at plant p at period n.

$$\sum_p St_{pin} - Back_{in} = \sum_p St_{pin-1} - Back_{in-1} + Pr_{in} - dem_{in} \quad \forall i, n \quad (14)$$

The raw materials mass balance is forced by eq.(15). The variable $Q_{pp'rn}$ denotes the quantity of raw material r transferred from plant p to plant p' at period n, while TR_{sprn} represents the amount of raw material r transferred from supplier s to plant p at period n. Parameter φ_{ri} is the stoichiometry coefficient of raw material r for producing final product i. St^{RM}_{pin} stands for the stocked amount of raw material r at plant p at period n.

$$St^{RM}_{prn} = St^{RM}_{prn-1} + \sum_{s \in (SR_r \cap SP_p)} TR^{RM}_{sprn} + \sum_{p' \in PP_p} Q_{p'prn} - \sum_{p' \in PP_p} Q_{pp'rn}$$

$$- \sum_{i \in IR_r} \sum_{j \in (JP_p \cap JI_i)} \phi_{ri} B_{pijn} \quad \forall p, r, n \quad (15)$$

3.6. Capacity constraints

Equations considering maximum storage capacity of raw materials and maximum storage capacity of products for every production facility, p, are also included. Moreover, supplier capacity restrictions, regarding the amount of each raw material type that can provide, are considered. Finally, transport capacity of raw materials between each pair of production facilities is also taken into account.

3.7. Objective function

The minimization of backlogs, inventory, sequence-dependent, setup, operating, raw materials' purchase/transportation and raw materials' interplant transportation costs are the optimization goal.

4. Case study

It follows a brief description of the main special features of this case study in order to demonstrate its complexity. Afterwards, the results obtained are thoroughly discussed.

4.1. Case study description

In this section, it is addressed the multi-site batching/scheduling and planning problem of 20 final products (*i1-i20*) which can be produced in 5 different processing unit types (*J1-J5*) of 2 different production plants (*P1-P2*). 6 different raw materials' types (*r1-r6*), which are provided from 4 different suppliers (S1-S4), are necessary to produce the set of final products. Products demand arrives in a weekly basis. 4 planning periods (168 hours each) have been considered. Production plants are not identical; production plant *P1* has unit types *J1* to *J5*, while production plant *P2* has installed units *J1* to *J3*. Suppliers *S1*, *S2* and *S4* can provide raw materials to *P1*, and suppliers *S1-S3* can provide raw materials to *P2*. Supplier *S1* provides *r1-r3*, S2 provides r2-r5, *S3* provides *r1* and *r6*, and *S4* provides *r4-r6*. 1 final product (*i20*) requires 3 different raw material types. 13 final products consume 2 different raw material types while 6 final products need 1 raw material type. 1 product (*i8*) can be done in 4 different units. 6 products can be produced into 3 different unit's types. 12 products can be processed in 2 different unit's types and 1 product (*i10*) can be produced just to 1 unit's type. Raw materials can be transferred between production facilities. The minimum production time is equal to 10 hours for all units but unit *J5* wherein it is equal to 5 hours.

4.2. Case study results

The model consists of 6,744 equations, 1,444 continuous variables and 5,256 binary variables. The problem has been solved to optimality in GAMS 22.8 (CPLEX 11.0), in a Dell Inspiron 1520 (2 GHz, 2 GB RAM), after 430 CPU seconds.

Fig.3 shows the contribution of each production facility to the total inventory profile for every planning period. There is no inventory at the last planning period. There are no backlogs. Fig.4 shows the multi-site Gantt chart for both plants at all planning periods. Raw materials flows are observed between the two production plants.
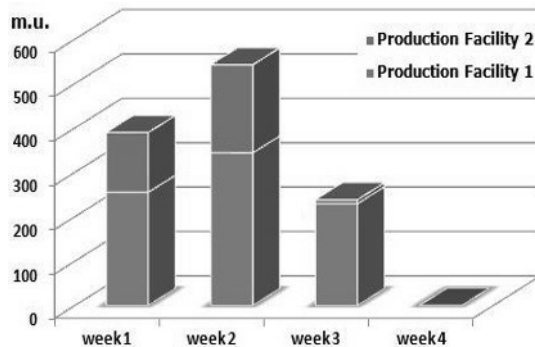
Figure 3. Inventory profile for the case study
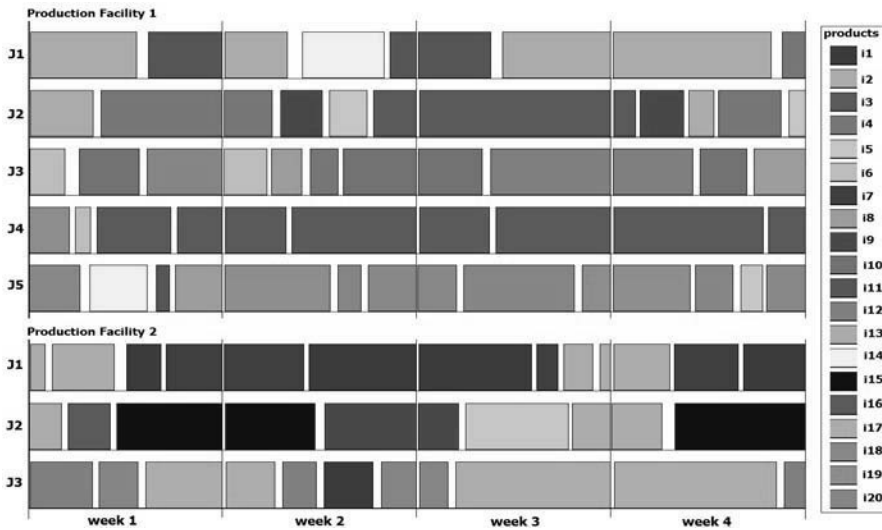
Figure 4. Gantt chart for the case study

5. Final considerations

In this work, a new multi-site batching/scheduling and planning MILP framework has been proposed. Raw material flows are allowed among production facilities; exploring the flexibility of these operations. Sequence-dependent costs are explicitly calculated avoiding the generation of ostensible optimal solutions. The ability of our approach to split sequence-(in)dependent times across the border of adjacent planning periods results to production capacity utilization improvement. Our current research activity is focused on the expansion of the current model to cope with more complicated product recipes, where intermediate products are present. All materials flows among the nodes of the supply chain entity will also be considered in order to take into advantage the flexibility of logistics operations. Finally, decomposition techniques will be revised and built up with the purpose of tackling large-scale industrial problems.

Acknowledgements

References

E. Alvarez, 2007, Multi-plant production scheduling in SMEs, Robotics and Computer-Integrated Manufacturing, 23, 608-613

M.H. Bassett, P. Dave, F.J. Doyle III, G.K. Kudva, J.F. Pekny, G.V. Reklaitis, S. Subrahmanyam, D.L Miller, M.G. Zentner, 1996, Perspectives on model based integration of process operations. Computers & Chemical Engineering, 20, 821-844

C.A., Floudas, X. Lin, 2004, Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review. Computers & Chemical Engineering, 28, 11, 2109–2129

N. Shah, 1998, Single and multisite planning and scheduling: current status and future challenges, Proceedings of the 3rd International Conference on FOCAPO, 75-90

C. Sung, C.T. Maravelias, 2008, A mixed-integer programming formulation for the general capacitated lot-sizing problem, Computers & Chemical Engineering, 32, 1-2, 244-259

Optimal Location Planning for Self-Storage Enterprises

Richard Lackes[a], Markus Siepermann[a]

[a]*Department of Business Information Management, Technische Universität Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany*

Abstract

The location planning problem is a long-term problem that aims for the determination of storehouse sites and the point in time when they should be put on stream. The decision criterion consists in the net present value. The paper introduces a decision model based on binary decision variables representing the sites and the points in time of investment with respect to a given budget and the constraints of the market situation and volume.

Keywords: Location planning, self-storage, decision model, genetic algorithm

1. Objective

In the past decade, an innovative concept for storehouses evolved with its origin in the USA (see Self Storage Association 2007). For outsourcing purposes, service providers offer storage capacities for individuals as well as for business users. The market is promising because the investment in storehouses, the operating costs and the market penetration are relatively low while the potential demand is high (see Duffy/Kliebenstein 2005). The basic idea is as follows: The service provider procures standardized storage room for a short period of time, the storage equipment (roller shutter, fork lift etc.) and administers the storehouse, but stockpiling and stock removal have to be done by the customers themselves. Due to an electronic entry system customers can access their rented storage capacities at every time independent of the presence of warehouse employees. For renters, this concept allows to reduce fixed storage costs that can now be replaced by usage dependent variable costs (see Mark 2005).

Because of the innovative character of the service and the developmentally chances many new sites will emerge in the next few years (see Duffy/Kliebenstein 2005). Therefore, it is very important in this stage to choose appropriate locations. Densely populated areas are attractive because of the restricted catchment area of a storehouse and the closeness to the potential target group. But the competitive situation and the investment costs in these areas are normally inauspicious. Thus, we are facing a complex long term site planning problem: How to choose the site that is the economically favorable one for the next years? (see Fleischmann/Klose 2004) In the following we present a multi periodic optimization model for this problem and show how this problem can be solved.

2. Development of the Optimization Model

2.1. Characteristics and Goal

The characteristics of self-storage storehouses (SSS) are (see Duffy/Kliebenstein 2005): (1) Construction and equipment of SSS are considered as a medium-/long-term investment. (2) Once a decision on investment and location is made, a revision can't be taken without greater loss. (3) SSS provide a certain capacity of storage space. (4) The offered

product »storage possibility« isn't affected by usage concerning its quality and life expectancy. (5) Operating costs of an SSS aren't constrained by use and load. They just ensure the disposability (availability fees). Thus, the marginal costs of an additional contract, if it lies within the capacity limits, matches 0. Non-use of storage capacity doesn't diminish the operating costs. (6) The sales market of an SSS is locally bounded to the location of choice. Main target group are individuals and craftsmen. (7) Even if the rental contracts allow flexible durations, most of the contracts are on a long term basis. To convince a customer once is important for the »natural« customer loyalty.

The location planning for self-storage enterprises is a multi-periodic dynamic decision problem. The planning horizon amounts to T years, scaled in t=1,2,...,T periods. Opening of storehouses takes place at period begin. The present point in time is t = 0. The goal is the maximization of the net present value that is determined by all site decisions made within the planning horizon. The decision concerns the expansion strategy. That is if and when storehouses should be built at a location within the planning horizon.

2.2. Optimization Model

2.2.1. Site Alternatives and their Characteristics

The investigation area shall be split into equal grid boxes (e.g. 5*5 km), each regarded as an »atomic« location element. A location will be determined by its x- and y-coordinates (x,y). Let x=1,…,X and y=1,…,Y be the relevant coordinates of all locations, so that the complete surface of the investigation area can be covered. Irrelevant locations that lie outside the grid because of the irregular shape of the investigation area are initially kept for an easier, formal description even if they fall apart later on. Each location is characterized by a set of attributes. Depending on their values a location rating can be computed. The relevant attributes of a location (x,y) for period t = 0,…,T are:

1. Outpayment: Land prices $GP_{x,y,t}$ [$]; storehouse equipment costs $LEQ_{x,y,t}$ [$]; labor costs index $LNIV_{x,y,t}$; annual outpayment-effective operating costs $K_{x,y,t}$ [$/period].
2. In-payment: Population $POP_{x,y,t}$ measured in 1.000; purchasing power of population, measured by purchasing power index $KKI_{x,y,t}$; market range of coverage, attainable price per square meter $P_{x,y,t}$ [$/sqm]; average rented storage space per contract [sqm]; life cycle curve of »salable« storage space (contracts or rather rented space) conditioned by age of the storehouse; competitive situation (foreign as well as one's own SSS in catchment area); economic trend

2.2.2. Prerequisites and Decision Variables

For the optimization, the following conditions shall hold: (V1) At each location should be built a maximum of one storehouse. Locations with an already existing storehouse aren't considered any further (see further V2). (V2) Shutting down of storehouses won't be allowed. (V3) There exist competitors on the market. (V4) Due to financial shortage or other bottlenecks only Bt storage houses can be built in one period t. (V5) Each storehouse provides a certain maximum capacity of storage space KAPx,y (e.g. 4200 sqm). (V6) The periods aren't subdivided any further. All payments, except acquisition payments, occur at the period-end.

The decision variables consist of binary variables differenced after the locations and the construction periods (see Vahrenkamp 2007). Because of condition V1, only the values 0 (no construction) and 1 (construction of an storehouse) can occur so that the optimization model is a binary decision problem with the decision variables $S_{x,y,t}$:

$$S_{x,y,t} = \begin{cases} 1 & \text{if a storehouse is built at (x,y) in t} \\ 0 & \text{else} \end{cases} \quad \text{with } x \in \{1,...,X\}, y \in \{1,...,Y\}, t \in \{1,...,T\}$$

At the start of planning, already existing storehouse locations are such $(x,y) \in X \times Y$ with $S_{x,y,0} = 1$. Because of condition V1 and V4 it applies formula (1) and (2):

(1) $$\sum_{t=0}^{T} S_{x,y,t} \leq 1 \text{ for all } (x,y) \in X \times Y$$

(2) $$\sum_{x=1}^{X} \sum_{y=1}^{Y} S_{x,y,t} \leq B_t \text{ for all } t=1,...,T$$

2.2.3. Determination of Outpayments for Equipment and Operation of a Storehouse

With regard to the outpayment, we have to consider site specific land prices $GP_{x,y,t}$ and site neutral payments for storage equipment $LEQ_{x,y,t}$ (e.g. 2 Mio \$). Furthermore, there occur operating costs which are almost fixed costs. With approximately 50%, labor costs are the biggest cost pool as surveys are showing. That means the annual site neutral costs affecting payments are K_t [\$/year] and the site specific costs – affected by the labor costs index $LNIV_{x,y,t}$ – are represented by $K_t \cdot LNIV_{x,y,t}$. The labor costs index $LNIV_{x,y,t}$ indicates the multiplier referring to a base salary (e.g. 1.07). From this, the annual costs affecting payments for the location (x,y) result in: $K_t \cdot (1 + LNIV_{x,y,t})$.

2.2.4. Determination of In-Payments

2.2.4.1. Calculation of Market Potential

A storehouse's market range of coverage is determined by its catchment area. It may reaches beyond its own location (x,y) and can also contain the ones nearby. Therefore, we define a degree of proximity $1 \geq Ng((x_1,y_1),(x_2,y_2)) \geq 0$ for all pairs of locations in such way, that they decrease with increasing distance from the observed location. It indicates which share of the population in (x_2,y_2) can be reached by a storehouse in (x_1,y_1) due to distance and transportation infrastructure. The degree of proximity of one's own location obviously is 1. Symmetry shall always apply. All locations in the neighbourhood with a positive degree of proximity are relevant for the site decision. With the help of this environment information the potential reachable customers $KUZ_{x,y,t}$ of a location $(x,y) \in X \times Y$ in period $t=1,...T$ is determined as:

(3) $$KUZ_{x,y,t} = \sum_{i=1}^{X} \sum_{j=1}^{Y} POP_{i,j,t} \cdot Ng((x,y),(i,j))$$

with $POP_{i,j,t}$ is the population of location grid box (i,j) in period t. If there are competing storehouses (own or foreign) that have access to the same market potential then the market potential has to be split. It has to be noted that not only storehouses of competitors but also own storehouses may reduce the market potential of a location (cannibalization effects). Let $L_{i,j,t}$ be the number of storehouses at the beginning of period t in location (i,j) without differencing of own and foreign storehouses. Thus, the starting situation is described by $L_{i,j,1}$ with $L_{i,j,1} \geq S_{i,j,0}$ because of the competing storehouses. Then, considering the starting situation and the site decisions within the planning horizon the value of $L_{i,j,t}$ $(t>1)$ can be computed as:

(4) $$L_{i,j,t} = L_{i,j,1} + \sum_{t'=1}^{t-1} S_{i,j,t'} \text{ with } (i,j) \in X \times Y; t=2,3,...,T$$

The »access intensity« $ZUG_{i,j,t}$ that describes how many customers can be reached in period t by the storehouse in location (i,j) can be defined as:

(5) $$ZUG_{i,j,t} = \sum_{k=1}^{X} \sum_{l=1}^{Y} Ng((k,l),(i,j)) \cdot L_{k,l,t} \text{ with } (i,j) \in X \times Y; t=1,2,...,T$$

Consequently, in period t the relevant market potential (in thousand inhabitants) of a storehouse to be built in location (x,y) is determined by:

(6) $MP_{x,y,t} = \sum\limits_{i=1}^{X}\sum\limits_{j=1}^{Y} POP_{i,j,t} \cdot \dfrac{Ng\big((x,y),(i,j)\big)}{\max\big\{Ng\big((x,y),(i,j)\big)+ZUG_{i,j,t};1\big\}}$

In a bottleneck situation, the market potential is distributed proportionally in accordance to the degree of proximity.

2.2.4.2. Attainable Price and Quantity of sales

Empirical studies have shown that the price per sqm $P_{x,y,t}$ correlates positively with the purchasing power index $KKI_{x,y,t}$. The capacity utilization depends except for the market potential $MP_{x,y,t}$ on the the age of a storehouse. Thus, there is a »life cycle curve« that can be described with the age dependent success rate $SUCCESS_s$ ($s=1,...,T_L$) measured in contracts per 1.000 reachable customers. T_L is the lifetime of a storehouse. In its beginning a storehouse becomes known and gets used until the capacity limit is reached. The typical curve is first ascending continuously up to a certain absorption point and then stagnating. Practical experiences have shown that this point usually is reached after six years. The reason for this phenomenon is that many customers are storing goods during a long period of time. Once a storehouse has gained a customer he most likely will rent his storage box over the next years. This effect is enhanced by relatively high costs for stock transfer if another storehouse will be chosen for rental. Therefore, the number of contracts $AK_{x,y,t}$ is determined by (t is the storehouse's building time):

(7) $AK_{x,y,\tau} \leq SUCCESS_{\tau-t+1} \cdot MP_{x,y,\tau} \qquad \tau = t, t+1,...T$

Additionally, the number of contracts is limited by the capacity of a storehouse. Let $LF_{x,y,t} = f_L(KKI_{x,y,t})$ with $dLF/dKKI_{x,y,t} \geq 0$ be the averaged storage space per contract. Then, the demand is determined by the number of contracts multiplied with the averaged storage space per contract.

2.2.4.3. The Investment's Residual Value

Let $RW_{x,y,t}$ be the residual value of a storehouse built in period t at location (x,y). This value represents a storehouse's value at the end of the planning horizon T. It is needed because the revenue of such an investment takes place after a certain period of time that might lie beyond the planning horizon. If $RW_{x,y,t}$ wouldn't be taken into account investments at the end of the planning horizon would be monetarily misinterpreted.

2.2.5. Objective Function

The acquisition value $AW_{x,y,t}$ of a storehouse at location (x,y) at the beginning of period t has two components: The site specific land price $GP_{x,y,t}$ as well as the site neutral payments for the storehouse equipment $LEQ_{x,y,t}$:

(8) $AW_{x,y,t} = GP_{x,y,t} + LEQ_{x,y,t}$

Then, the net present value $CV_{x,y,t}$ of a storehouse in location (x,y) built at the beginning of period t ($x \in X$, $y \in Y$, und $t=1,...,T$) will be:

$$CV_{x,y,t} = -AW_{x,y,t} \cdot (1+i)^{-(t-1)} + RW_{x,y,t} \cdot (1+i)^{-T} +$$

(9)

$$\sum_{\tau=t}^{T}\Big(AK_{x,y,\tau} \cdot LF_{x,y,\tau} \cdot P_{x,y,\tau} - K_{x,y,\tau} \cdot \big(1+LNIV_{x,y,\tau}\big)\Big)\cdot(1+i)^{-\tau}$$

The discounting always is for the planning horizon begin, i.e. t=0. The acquisition payments accrue at period begin, all other payments at period-end. Naturally, the following condition must hold:

(10) $CV_{x,y,t} \geq 0$ with $(x,y) \in X \times Y$; $t=1,2,...,T$

Now, the objective function consists in the maximization of the total net present value:

(Z) $\quad$ Max $\displaystyle\sum_{t=1}^{T}\sum_{k=1}^{X}\sum_{l=1}^{Y} CV_{x,y,t} \cdot S_{x,y,t}$

Because $S_{x,y,t}$ are binary variables we are facing a binary decision model. The number of decision variables is $X \cdot Y \cdot T$. In order to compute the access intensity $ZUG_{i,j,t}$ all existent storehouses in t plus those storehouses to be built (represented by $S_{x,y,t}$) including the proximity index have to be considered. Additionally, the access intensity is a determination factor of the market potential. Thus, the decision model is NP-hard and cannot be solved in an acceptable calculation period. Even commonly known optimization algorithms like e.g. branch and bound have to compute all solutions in order to find the optimal solution. Only with proximity index 0 for all neighbour grid boxes a classical binary optimization algorithm could succeed because then, the access intensity only depends on the known starting situation at the beginning of the planning horizon.

3. Example

Let us now apply the presented model to Germany. Figure 1 shows a map of Germany on the left side. On the right side we can see the grid boxes the investigation area is split into. The grey grid boxes indicate the six areas with the highest population density: Hamburg, Berlin, Munich, Frankfurt, Cologne and the Ruhr. These are the regions the model is coping firstly. Due to the relatively low outpayments in Berlin and the Ruhr, storehouses are first of all built in respectively beside these regions. The chosen locations depend on the values we are using for each grid box and the access intensity. If the outpayments of the white boxes are low in comparison to the grey boxes and the access intensity is not zero the chosen locations lie outside the six areas. Otherwise, if each region consists homogenously of boxes with equal values the chosen locations lie in the center of each region.
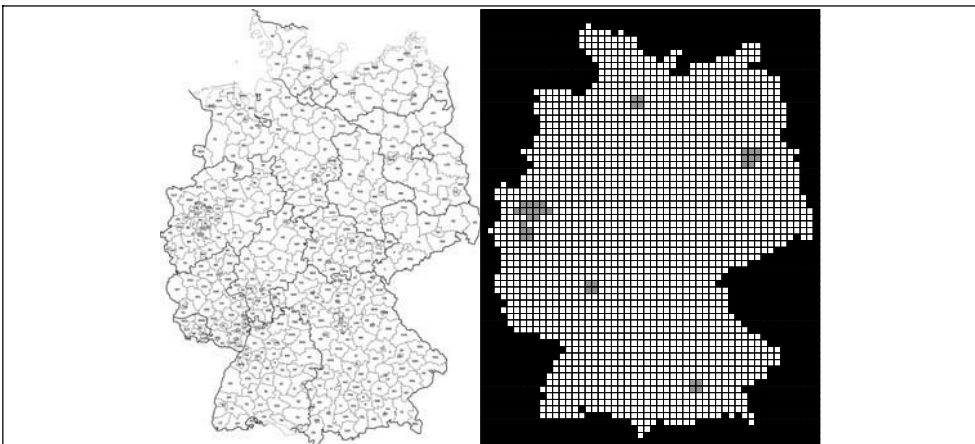
Figure 1: Investigation area Germany

The calculation for Germany could only be done because we reduced the number of grid boxes without empty values and the number of periods so that the number of possible solutions was decreased enormously. If we use a totally filled grid the calculation period explodes: With five periods and a 100×100 grid the solution space consists of 9,5 billion elements that have to be computed. Therefore, we will use a genetic algorithm in order to find good solutions in an acceptable time.

4. A Genetic Algorithm as Approach

4.1. Individuals

In the following we assume $B_t = 1$ for simplification. A solution can be described via a 3D-cube with the location coordinates and the periods as dimensions (in the following see Mitchell 1998, Vose 1999). Then, an individual of the genetic algorithm is one solution alternative that can be defined as follows: The individuals can be represented as a X×Y-matrix M. The matrix contains a maximum of T values between 1 and T whereas no value occurs twice. The other values of the matrix are 0. A value $M_{x,y} > 0$ in the matrix indicates that a new storehouse is built at site (x,y) in period $M_{x,y}$. The value 0 indicates that no storehouse is built at site (x,y). The net present value is used as fitness to choose the individuals for crossover and the selection of the next generation.

4.2. Mutation

The mutation operator can be defined as follows: Randomly choose two cells of an individual M with different values and swap the two values. Because of the crossover we are discussing later on some values might be lost during the calculation. In order to reproduce those missing values between 1 and T we can insert them instead of swapping two values.

4.3. Crossover

Let M and N be individuals of the genetic algorithm. Then, the crossover operator exchanges parts of the two individuals as follows: Randomly choose the bounds of a rectangle in the matrix. Then the rectangles are cut out of M and N and implanted into the other solution. This operation may lead to invalid individuals: The number of values between 1 and T may now be greater than T (case 1). Values between 1 and T may occur twice but there are only T values greater than 0 (case 2). Therefore, a repair mechanism has to be installed. In the first case, we randomly choose a cell with a value that occurs twice and set the value of the cell to 0 until there are only T values greater than 0. In the second case there are doublets as well as missing values. Therefore, we randomly choose a cell with a doublet value and set the cell's value to one of the missing value until no value occurs twice.

5. Conclusion and further Enhancements

In this paper we presented an optimization model for the location planning of self-storage enterprises concerning the expansion strategy. As this problem is a binary decision problem with many decision variables it can hardly be solved with deterministic algorithms. As genetic algorithms rapidly find good solutions (see Koza 1993) we designed a genetic algorithm that finds a good solution in an acceptable computation time.

References

Duffy, S., Kliebenstein, R.K., 2005, How to invest in Self-Storage, MiniCo Publishing.

Fleischmann, B., Klose, A. (eds.) 2004, Distribution Logistics: Advanced Solutions to Practical Problems, Springer.

Koza, J.R., 1993, Genetic Programming, MIT Press.

Mark, K., 2005, Leveraging Wal-Mart, eBay and USPS, in Morse, E. A., Mitchell, R. K., Cases in Entrepreneurship: The Venture Creation Process, SAGE.

Mitchell, M., 1998, An Introduction to Genetic Algorithms, MIT Press.

Self Storage Association, 2007, US Self Storage Industry 1985 to 2005: Facilities, Square Meters, Revenue etc., Self-Storage-Conference 2007, Paris.

Vahrenkamp, R., 2007, Logistik, 6. Auflage, Oldenbourg.

Vose, M.D., 1999, The Simple Genetic Algorithm: Foundations and Theory, MIT Press.

What, if Anything, is a Chemical Engineer?

Laureano Jiménez-Esteller, Gonzalo Guillén-Gosálbez, Dieter T. Boer

*Department of Chemical Engineering, School of Chemical Engineering, University
Rovira i Virgili, Av. dels Països Catalans, 26, Tarragona, 43007, Spain. E-mail:
{Laureano.Jimenez,Gonzalo.Guillen, Dieter.Boer}@urv.cat*

Abstract

The core domains of Chemical Engineering are in crisis as corresponds to a mature
discipline that has been going on for more than 100 years. However, over the last two
decades, some concerns have been raised about the identity and future of Chemical
Engineering and our ability to adapt to new challenges: (1) Chemical engineering (ChE)
graduates accept positions in an extremely vast a diverse range of sector, all having
specific knowledge requirements; (2) globalization of the chemical industry forces
professionals to work with people with different values and principles; (3) the research
scope of ChE has expanded, and now the *hot areas* are at the interfaces with other
disciplines, more focused in products than in processes, and covering the spectrum from
the macroscopic to the molecular level. To adapt to this new scenario, an educational
model was designed and implemented at the School of Chemical Engineering (ETSEQ)
in Tarragona (Spain) to enable ChE students to acquire and integrate technical and
scientific knowledge through the simultaneous and gradual development of
competencies encompassing social and management skills. This model is based on the
large-scale deployment of a project-based cooperative learning approach throughout the
curriculum.

Keywords: problem-based solving, non-technical skills, project based learning.

1. The realm of ChE is large and diverse

I would like to begin this communication by explaining its title. I have borrowed it from
the naturalist Stephen Jay Gould [Gould, 1984]. Gould wrote an essay called '*What, if
anything, is a zebra?*' which objective was to answer if zebras are white animals with
black stripes or black animals with white stripes. The beauty of this title (originally
from A. E. Wood, 1957) is that it allows to discuss something that is clear in all our
minds (identity of Chemical Engineering), but which role and functions are
continuously evolving to solve old problems, to answer new questions, and preparing
the topics that will have to elucidate in the (uncertain) future.

Coming back to chemical engineering, it is commonly accepted that ChE has progressed
through three different paradigms that promoted an equal number of metamorphosis:

1. The concept of unit operations, developed by Arthur Little ($\approx$ 1910) as a notion to
   understand the underlying processes common to all chemical products, reactions, and
   machinery.
2. The inclusion of transport phenomena, thermodynamics and kinetics in the design of
   unit operations and the unified mathematical description of mass, energy, and
   momentum transfer [Bird *et al.*, 1960].
3. The extension of the domain to the interfaces with other disciplines, like
   biotechnology, medicine, food technology or environmental sciences. As a proof of

its impact, during the last two decades an increasing number of Chemical Engineering Departments have changed their names and now their affiliations include other areas, like "*environmental*" or "*biotechnology*".

1.1. Is chemical engineering in crisis?

Following with the evolutionary analogy of the title of this work, the transition of each one of the paradigms was a time for crisis, where different, and sometimes conflicting approaches, were explored (similar to the extinction periods of the evolution). Right after each new paradigms was established, a period of great richness and exciting findings in the new directions flourished (like the explosion of biological patterns and *phyla* that followed each extinction periods).

As time progresses, the technology we have been developing during the last decades give the impression to be incompatible with the problems that seems to arise (*e.g.*, global warming, water use, energy production…). In this context, the spatial scale of chemical engineering studies could range from nanotechnology (polymers, composites, electronics, coatings, catalysis, drug discovery, diagnostics…) to huge macrosystems (*e.g.* the boundaries of the system must include all upstream and downstream activities to study the environmental and sustainability problems). On the other hand, the temporal scale can cover from nanoseconds (nuclear reactions, particle interactions…), to centuries (*e.g.* stability of radioactive waste disposals, long-time effects of chemicals in the environmental and health…).

The need for change is evident and we must face those challenges. Sustainability and green technologies probably will be the cornerstone of chemical engineering of the future. The energy demand is rapidly increasing and the society will look at engineers to find a solution. One of our tasks is to prepare the technology to deal with the important issues that will be derived from the peak-oil, a very sensitive topic in geopolitics. At a given time, we will be forced to change from carbon-based energies to more sustainable sources. Technologies that now represent a small percentage of the power generation (wind, solar, hydro electrical, oceans, bio-fuels, geothermal, nuclear…) are expected to growth and should reach their maximum potential very quickly with a robustness similar to the actual resources. We can have a key role in the design, implementation of large scale facilities and improvements of these technologies.

If we consider the '*Big Four*' engineering fields (civil, mechanical, electrical, and chemical), our domain is numerically the smallest. However, on average, chemical engineers consistently top the list of the higher starting salary with a bachelor degree in the USA [NAS, 2007]. This issue highlights that, at least from the demand side, there is not such a crisis.

2. Benchmarking of chemical engineering

The work by Marrero [Marrero, 2008] focuses on obtaining data about the number of ChE graduates and their present and future employment on a world-wide basis. Table 1 indicates the increase of places to study Chemical Engineering. The study indicates imbalances between supply and demand sides. In addition the flow of professionals from one nation to another is affected by political constraints.

2.1. Comparison of chemical engineering undergraduate programs by multivariate statistical analysis

This study compares ChE undergraduate programs from all over the world ($\approx$120 Universities) to obtain patterns based on the location and curricula. The work is based on the information available on the internet, and thus the bottleneck was the lack of specific course descriptions.

Table 1. Number of nations and schools with Chemical Engineering [Marrero, 2008].

	North America	South America	Africa	Asia	Europe	Mid-East
	# of nations					
1985	3	7	4	12	21	2
1995	3	5	4	11	21	5
2005	3	5	1	12	20	6
	# of schools					
1985	177	24	10	93	110	4
1995	181	20	10	79	116	9
2005	179	14	5	64	88	9

The ChE curricula were classified in 24 categories, divided in practical and theoretical classes. Multivariate statistical techniques (cluster analysis, linear principal component analysys and linear discriminant analysis) were applied to study the similarities between undergraduate programs. Data were pre-processed to use the percentages dedicated to each category, thus comparing the structure of the curriculum, not the duration [Jiménez and Mateo, 2003].

The average percentage of theoretical classes is around 55%, an average value for disciplines as ChE. Results from the principal component analysis show that most of the 24 categories do not group according to any predefined pattern, except mass and energy unit operations and mathematics and physics. When practical hours are considered, the cluster analysis shows sub-clusters, and there seems to be a certain similarity between Spanish speaking universities. In fact, most of the variance of the curricula depends on the practical classes, while theory does not have such a rich diversity.

The discriminant analysis shows that universities are not randomly distributed. Results show that over 80% of them were correctly allocated by the model to the geographical region they belong to (if the allocation were completely random, we would only expect a 14% correct allocation). The authors can not state if the similarities are due to cultural reasons rather than idiomatic (*e. g.*, USA, Australia). Also, the homogenization effect of accreditation programs (*e. g.*, ABET or Bologna process) may have some impact.

3. Approach at the School of Chemical Engineering

An educational model has been implemented at the School of Chemical Engineering (ETSEQ) at the University Rovira i Virgili (Tarragona, Spain), to enable ChE students to acquire and integrate technical knowledge encompassing social skills (teamwork, cooperation, planning, decision making, problem-solving abilities, communication skills…). In all courses, classical teaching, experimental and virtual laboratories and computer aided process engineering tools had to be coordinated to solve an open-ended project (*i. e.* with many valid approaches and different proper solutions).

3.1. Does faculty research improve undergraduate teaching?

Faculty members and educational researchers have been arguing for decades whether research supports undergraduate training. Those who agree with this statement cite many ways in which research can enrich teaching *a priori*, while those on the other side state that all studies consistently fail to show any measurable linkage between both activities [Prince *et al.*, 2007].
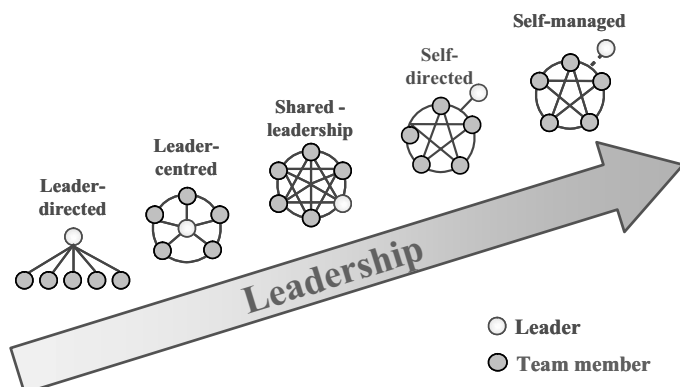
Figure 1. Team-work development [Witt *et al.*, 2006].

At the School of Chemical Engineering (ETSEQ), the research-teaching *nexus* have been exploited through two specific strategies: bringing research into the classroom (and to the society, through popular science) and involving undergraduates in research projects ('*Research laboratory*', a 75 hours compulsory subject in the 9th or 10th semester). The first topic is not covered by faculty members with research projects bringing their results into the classroom, as research contents is well over the level of most undergraduate students. To better realize the potential synergies between faculty research and undergraduate education, students are introduced into a research group to develop a specific task (typically around 40 students per year course this subject, a third of them in companies). During this period, students are exposed to solve routinely open-ended, ill-posed and ill-defined problems, figure out what they need to know, search scientific sources to obtain the missing information, present an intermediate report with the hypothesis, propose some test to find possible solutions, obtain the final results, explain them in a written report and defend them in an oral presentation.

Faculty members act as mere advisors and consultants so students are not subjected to passive teaching [J. Ziemlewski, 2009]. Undergraduate students are actively involved, enhance understanding, improve retention, become proficient in problem solving, self-directed learning, build decision criteria and team participation (cooperate rather than compete). Obviously, we do not expect students magically develop their entire individual potential within this subject but we see some progress (Figure 1).

3.2. 2. Project-based learning and teaching methodology

The educational model followed at the ETSEQ is based on the large-scale deployment of a project-based cooperative learning approach throughout the ChE curriculum: 1st, 2nd, 3rd, 4th and 5th courses [Witt *et al.*, 2006; Giralt *et al.*, 1994a, 1994b; Jiménez *et al.*, 2002, 2003].

The expertise in team development and change management methodologies, gained from nearly a decade of implementation work in Tarragona, has complemented the practice of the School of Chemical Engineering with empirical and pragmatic learning methodologies. A set of external training interventions has been designed to support the development of competencies by students. In this educational approach student teams grow from leader-directed teams in the first semester of the first academic year to self-directed (or empowered) teams in the fifth year. In this empowerment journey, fourth-year students play a key role as they act as facilitative leaders of first-year and second-year project teams, adjusting their leadership role according to the team development stage (Figure 1).

For example, two eight hours modules are taught during the first two weeks of each term to first-year students. The modules are structured to take into account the factors needed for students to understand and commit themselves to change. The first-year students' evaluation shows that the module helps them:

- To identify what they need to accomplish to gain future employment as Chemical Engineers, thus settling the patterns to become life-long learners.
- To understand, from a practical point of view, what an integrated design project consists of and what the benefits of teamwork are.
- To realize that the integrated project and the related teamwork are great opportunities to acquire competencies that are essential in today's workplace.

In this way, hands-on teamwork training is implemented through specific seminars. These modules deal with different teamwork-related issues such as, leadership, team capabilities, common purpose, team norms, communication, conflicts, team operating procedures, member integration, team evaluation… These activities are distributed across the curricula, considering a long-term deployment. The activities are organized with optional attendance; nevertheless our experience is that more than 80% of students follow these modules, which can be validated as official credits (*i. e.* elective courses). In each session, two instructors that play different roles are available: one leads the session while the other acts as facilitator.

The core of the competency-based educational model is client orientation. The need to satisfy clients and to adapt to their changing needs triggers the development of competencies related to the transformation of the students (versatility, entrepreneurship and innovation...), of the organization (facilitative leadership, teamwork, cooperation…), and of the institution (organizational development, organizational performance, organizational leadership…). For a detailed distribution and allocation of each capability in the undergraduate ChE program, check the team development matrix that is explained in detail in Table 2.

Table 2. How to assess the team-work capabilities: team development matrix deployment over the ChE curriculum [Witt *et al.*, 2006].

Responsibility for work activity (Leader / Shared / Team members)	Leader directed	Leader centered	Shared leadership	Self-directed	Self-managed
1. Formulate team goals	Leader	Shared	Team members	Team members	Team members
2. Communicate outside the team	Leader	Shared	Team members	Team members	Team members
3. Decide on team rules	Leader	Shared	Team members	Team members	Team members
4. Resolve conflicts within the team	Leader	Shared	Team members	Team members	Team members
5. Solve problems	Leader	Shared	Team members	Team members	Team members
6. Interact with customers	Leader	Shared	Team members	Team members	Team members
7. Planning work, meetings, tasks, deliverables, etc.	Leader	Shared	Team members	Team members	Team members
8. Provide feedback to other team members	Leader	Leader	Team members	Team members	Team members
9. Improve/define work processes	Leader	Leader	Shared	Team members	Team members
10. Allocate work assignments	Leader	Leader	Shared	Team members	Team members
11. Plan training and development for team and members	Leader	Leader	Shared	Team members	Team members
12. Allocate resources within the budget	Leader	Leader	Shared	Team members	Team members
13. Determine recognition and rewards	Leader	Leader	Shared	Team members	Team members
14. Measure team's progress towards the goals	Leader	Leader	Shared	Team members	Team members
15. Decide on leadership within the team	Leader	Leader	Shared	Team members	Team members
16. Monitor expenses to be within a budget	Leader	Leader	Leader	Team members	Team members
17. Select new team members	Leader	Leader	Leader	Leader	Team members
18. Evaluate performance	Leader	Leader	Leader	Leader	Team members
19. Distribute and allocate scores	Leader	Leader	Leader	Leader	Team members
20. Integrate and train new members	Leader	Leader	Shared	Leader	Team members

4. Conclusions

The training in non-technical capabilities helps our students to set the pattern to become successful life-long learners. Student's main objections to the project-based cooperative learning approach were the excessive time devoted to the project. In addition they demand more effort in supervising, in particular in the initial steps of the project development. We realize that, as students are not used to this kind of teaching, at the beginning of the project more continuous help and guidance is needed than in traditional teaching methodologies. However, the lack of information forces students to use their own initiative to solve the open-ended problems (*i. e.* there are many valid approaches and solutions). For example, the inclusion of computer aided process engineering tools (*i. e.* virtual laboratories) help students to acquire the insights of the unit operations, since mathematical models are not useful for a qualitative interpretation of how design variables influences the unit operation performance.

Preliminary results show that student attendance has increased, that drop out has decreased, that more professors act as facilitators, and that active-oriented and student-centered educational methodologies are increasingly applied. The academic staff also needs to improve our performance as a teamwork model role for students, as sometimes we are surpassed by everyday research tasks and teaching is not in the top list of priorities. To sum-up, our opinion is that the benefits of applying this methodology largely exceed the tremendous effort required.

References

R. B. Bird, W. E. Stewart and E. N. Lightfoot, 1960, Transport phenomena. J. Wiley and Sons, New York, USA.

F. Giralt, M. Medir, H. Thier and F. X. Grau, 1994a, A holistic approach to the ChE education. Part 1. Professional and issue-oriented approach. Chemical Engineering Education, 28, 122-127.

F. Giralt, A. Fabregat, X. Farriol, F. X. Grau, J. Giralt and M. Medir, 1994b, A holistic approach to the ChE education. Part 2. approach at the introductory level. Chem. Eng. Ed., 28, 204-213.

S. J. Gould, 1984, What, if anything, is a zebra? Essay in Hen's teeth and horse's toes. Norton Paperback, New York, USA.

L. Jiménez, J. Font and X. Farriol, 2003, Unit operations laboratory using ill-posed problems, International Journal of Engineering Education, 19 (5), 717-720.

L. Jiménez and J. M. Mateo, 2003, Comparison and characterization of chemical engineering undergraduate programs by multivariate analysis, European Congress on Chemical Engineering, Granada, Spain.

L. Jiménez, J. Font, J. Bonet and X. Farriol, 2002, A holistic unit operations laboratory, Chemical Engineering Education, 36, 150-155.

National Academy of Sciences, 2007, International benchmarking of U.S. chemical engineering research competitiveness in brief. Availble at (April 2009): http://dels.nas.edu/dels/rpt_briefs/benchmarking_chem_eng_brief_final.pdf.

T. R. Marrero, 2008, Global chemical engineering workforce statistics, AIChE anual meeting, Philadelphia, USA.

M. J. Prince, R. M. Felder, R. Brent, 2007, Does faculty research improve undergraduate teaching? An analysis of existing and potential synergies, Journal of Engineering Education, 96 (4), 283-294.

H. J. Witt, J. R. Alabart, F. Giralt, J. Herrero, L. Vernis and M. Medir, 2006, A competency-based educational model in a chemical engineering school. Internationl Journal of Engineering Education, 22 (2), 218-235.

A. E. Wood, 1957, What, if anything, is a rabbit? Evolution, 11, 417-425.

J. Ziemlewski, 2009, Designing the new global chemical engineer, Chemical Engineering Progress, 105 (2), 6-10.

Information Modelling: Industrial Standards for Integrated Plant Management

Zofia Lukszo

Department of Energy and Industry, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, the Netherlands

Abstract

The paper describes the Master course "Integrated Plant Management" developed for the students of the Faculty Technology, Policy and Management at the Delft University of Technology. During the regular classes students learn advanced analytical methods for problem solving of long and short-term planning, forecasting and scheduling problems as well as process optimization. During the instructions they make an IDEF0 model of an industrial plant according to the standards ISA88 and ISA95. In this model the same activities as discussed during regular classes are placed in the context of manufacturing execution systems. This helps students in understanding the management complexity of an industrial plant and in contributing adequately to implement many challenging simulation and/or optimization solutions in real-life cases.

Keywords: Integrated Plant Management, Information Modelling, IDEF0, Industrial standards

1. Introduction

Process system engineering research into methods to support an (eco-) efficient process design and operation is gaining momentum since many years. Many contributions to the PSE research area concentrate on a specific aspect of the process design and operation: modelling and optimizing a process, plant or even an enterprise, defining more effective planning and scheduling algorithms, designing more advanced controllers etc. Some of these contributions have a strong theoretical character and therefore their practical applicability is rather limited.

Other contributions, often related to real-world industrial problems, also seldom find a way to be implemented in the industry and to contribute to the improvement of the overall plant performance. It is striking that the organizational aspects and the applicability of the valuable theoretical results are hardly explored. Academic courses aimed at providing knowledge and understanding of a plant operation in such a way that the challenges imposed by the economic, environmental and social sustainability are made more transparent, may change this situation. We strongly believe that scientific methods for first-principle modelling, rigorous optimisation and advanced control to be applicable in the industrial practice should be supported by a clear picture of the whole enterprise and working processes taking place there.

This paper contributes to the area of the integrated plant management. It describes the Master course "Integrated Plant Management" developed for the students of the Faculty Technology, Policy and Management at the Delft University of Technology. During the regular classes students learn advanced analytical methods for problem solving of long and short-term planning, forecasting and scheduling problems as well as process optimization (Edgar, 2001; Pinedo 2001; Grossmann, 2002; Biegler 2004). During the instructions they make an IDEF0 (Integration DEFinition of function modelling) model

of an industrial plant according to the industrial standards ISA88 and ISA95 (ISA, 1995; ANSI 2000, 2001, 2005). The activities as discussed during regular classes, e.g. scheduling, recipe optimization etc are placed during the modelling assignment in the context of manufacturing execution systems. At the end of the course an activity model "run a plant" is made and the students know who is involved in operational activities, where the necessary information is coming from, what is the interaction between these activities and to what extent changes at one system level contribute to the overall plant performance. This helps students in understanding the management complexity of an industrial plant and in contributing adequately to implementation of many challenging simulation and/or optimization solutions in real-life cases.

Section 2 introduces industrial standards ISA88 and ISA95 aimed at supporting operations management. In section 3, the activity modelling approach IDEF0 is described, followed by an illustrative example of activity decomposition for planning, scheduling and control functions. Finally, Section 4 gives some concluding remarks.

2. Industrial Standards for Integrated Plant Management

In the process industries there are very large financial incentives for fundamental business processes, e.g. planning, scheduling and control activities, to function in a coordinated and integrated fashion (Shobrys, 2002). Nevertheless, many industrial companies have not achieved desired integration in spite of multiple initiatives. Operating in a rapidly changing world of global economy and more short-term dynamics they continuously search for opportunities to improve their competitive position: the production processes produce products more efficiently and the internal methods of operation enable to be more effective. The awareness of the importance of environmental affairs and objectives are also set to the agenda of many enterprises. Mostly, during the improvement projects the companies concentrate on improving one task without taking sufficiently into account interactions with other activities and with the surroundings. Moreover, the formulated company objectives and policy are very often communicated inside the organization insufficiently. The language the people talk is different: top managers talk about profit and continuity, process people about quality and quantity and the environmental department about emissions and waste.

To contribute to the integrated plant management the modern PSE concepts should be properly placed in the total manufacturing execution system, so that it will be clear who is responsible in the plant for the particular activity, what information and from whom is needed to perform the activity, which other activity and to what extent depends on the generated results, and what is a possible contribution of the activity to the overall plant performance.

The ISA88 and ISA95 standards, published by the Instrumentation, Systems, and Automation Society, provide standard terminology and models for design as well as operation of (batch) control systems and operations management activities (ISA, 1995; ANSI 2001, ANSI 2005). Figure 1 presents the scope of both standards. The aim of the ISA88 standard is to provide a modular framework for recipe development and management. The aim of the ISA95 standard is to improve information exchange by providing a framework for the integration of the enterprise functions as strategic and tactical forecasting and planning, scheduling, quality assurance, maintenance management, procurement etc with the control and process execution layer as described by ISA88.
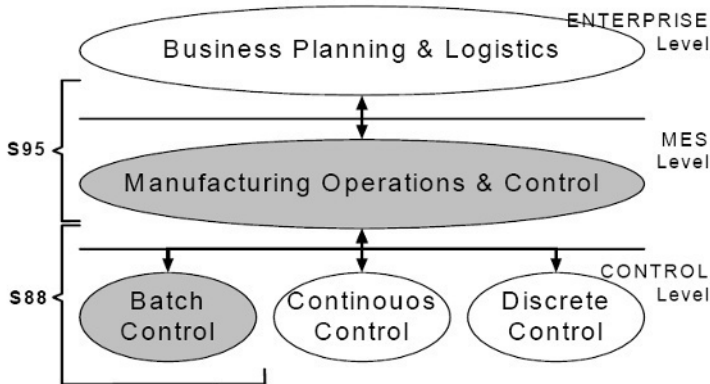
Figure 1. Scope of the ISA-88 and ISA-95 standards (grey areas)

The activities of manufacturing execution system (MES) layer are those activities of a production facility that coordinate the personnel, equipment, material, and energy in the conversion of raw materials and/or parts into products. Operations management including activities that may be performed by physical equipment, human effort, and information systems, corresponds to the activity model defined in the Part 1 of the ISA 95 standard. It defines 10 generic functions in an enterprise-control model and two additional functions as presented in Figure 2. Further, it describes point by point the tasks of each function. It should be stressed, that the model does not say anything about the importance and complexity of the functions. For example, material control for automotive industry is a very complex matter comparing to a fairly simple task in a water distribution company (Scholten, 2007).
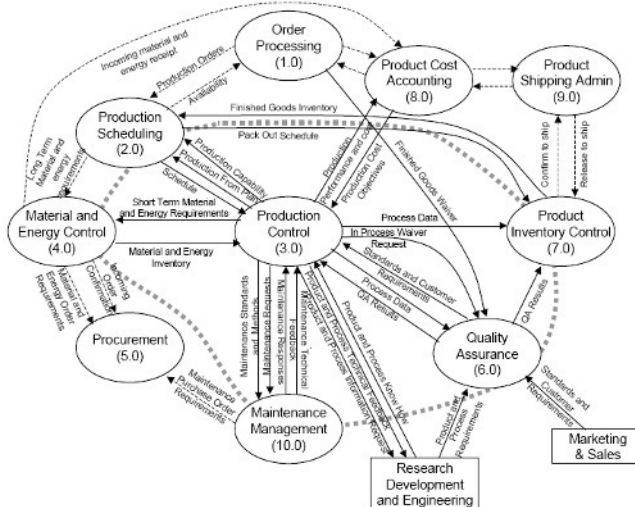
Figure 2. Functional ISA – 95 enterprise-control model

The bold dotted line in Figure 2 illustrates the boundary of the enterprise-control interface between the Enterprise Level and the MES Level (Manufacturing Execution System). The function *Production Scheduling* is placed on this boundary. At the Enterprise Level it generates production plans with a time horizon of months or weeks: master production schedule. On MES Level, detailed scheduling takes place based on days or hours, taking into account product demand pattern, available equipment, storage capacity and personnel, recipes, the relationship between cleaning time and product combination, available utility profile etc. The activity *Production Control* is a pure Control Level function. Every function exchanges information with other functions. ISA-95 groups this information into information flows; each with its own name and content, such as Schedule, Production Performance and Cost, QA results etc.

For the students is even important to understand how to perform scheduling activities by optimizing one or more adequate criteria (Mendez, 2006) resulting in an optimal schedule as to know all relevant information flows to create their own view on integrated plant management.

3. Key Business Processes and Their Mutual Relationships

Plant modelling according to the ISA-S88 and ISA-S95 standards by using the so called SADT (Structured Analysis and Design Technique) techniques and especially IDEF0 diagrams is very useful to support the understanding of integrated plant management. IDEF0 (Integration DEFinition of function modelling) offers a well-structured method for activity and information-flow analysis, focussing on the hierarchical decomposition of activities as well as on the interaction flows between activities (IDEF0, 1993). The primary objectives of the IDEF0 modelling technique are (Heijnen, 2006, Fuchnino, 2008):

- To provide means for complete and consistent hierarchical modelling of activities that take place in a system, and the data and objects that inter-relate those activities.
- To provide a modelling language that is generic (applicable for analysing systems and subject areas of varying purpose, scope and complexity); rigorous and precise (for producing correct, usable models) and concise (to facilitate understanding, communication, consensus and validation).

The whole system – in this case an industrial plant - is modelled in a clear way (there are only two symbols used: boxes and arrows) as a set of interrelated and nested activities. The first activity to be designed is the activity that describes the system itself, i.e. run a plant, and is called the context activity, see Figure 3. In IDEF0 not only the inputs, which are transformed into the outputs, are modelled, but also the controls and mechanisms. Controls (are fed to the top of an activity) are the objects that govern the manner in which inputs are transformed; yet they are not themselves transformed by the activity, e.g. in Figure 3 Market Information. Mechanisms (are fed to the bottom of an activity) are those objects that actually perform the transformation of inputs to outputs, yet they are not themselves transformed by the activity, e.g. software tools and methods, equipment and people. ICOM is an acronym for the categories of information presented in IDEF0 diagrams and represents four types of arrows: Inputs, Controls, Outputs and Mechanisms. The inputs are defined as the representation of the material or information that is consumed or transformed by the activity to produce outputs. Controls influence the activity without being transformed or consumed, they are necessary to perform the activity; they are often in the form of policies or procedures, and we extend them with the objectives. Outputs are the materials or information produced by an activity. Mechanisms are the resources that help to perform it. An output from one activity can

be an input to other activity. In this way, activities can be combined in a chain or a network.
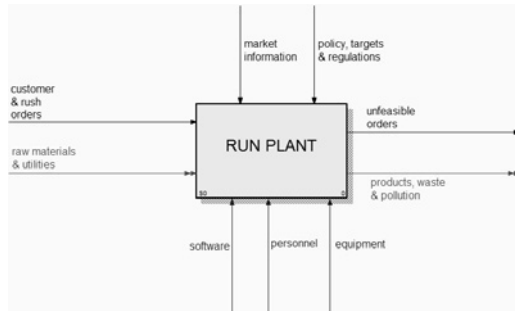
Figure 3. Context diagram 'Run a Plant'. Software: AllFusion Process Modeller from Computer Associates International

The context activity, as presented in Figure 3, defined at the highest level can be decomposed in a number of activities, which collectively should achieve the main objective of the plant. In this case the top activity "Run a Plant" can be decomposed further, e.g. into "Do Order Processing and Procurement", "Do Production Planning, Scheduling and Control, "Do Inventory Management", "Perform Product Cost Accounting", "Do Product Shipping Administration", "Do Quality Assurance". Next, each activity can be decomposed further, e.g. "Do Production Planning, Scheduling and Control". In the decomposed model an output from one activity can be an input to another activity. The feedback loops are easily understood. In this way activities can be combined in a chain or network.
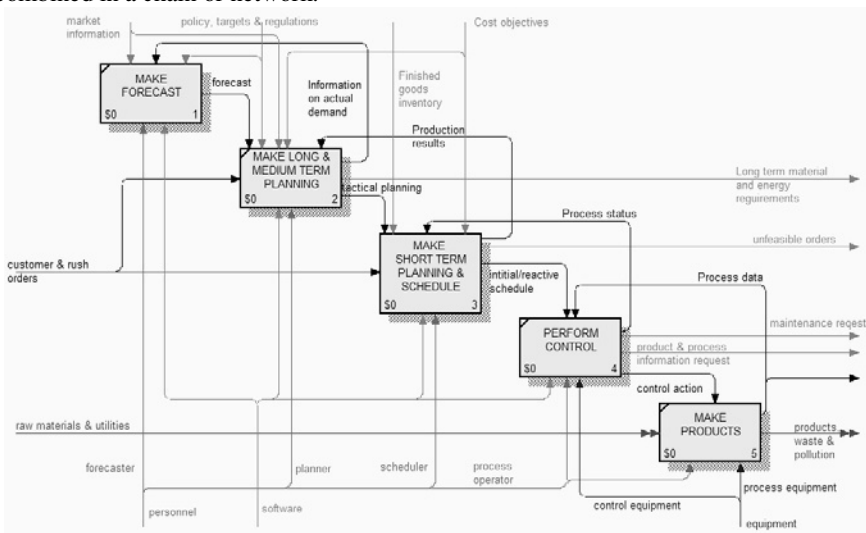
Figure 4. Decomposition of "Do Planning Scheduling and Control"

4. Final Remarks

Modelling enterprise activities and their interdependencies, and relating them to the operational objectives makes it possible to visualize how improvements or changes in

one activity interact with other activities and which results could be expected. This visualization supports the decision, which activities at the operational level contribute mostly to the overall objective of the plant, in other words which ones are most effective. Improvements in the efficiency of those activities will be most effective.

In an ongoing process together with partners from the industry, the students apply this method successfully to industrial projects (Salverda, 2006; Klompenhouwer, 2007, Stekelenburg, 2008). Through many industrial projects, effectiveness of the proposed procedure is confirmed. Relaying on these results, the activity modelling approach is shown promising in the implementation phase of PSE solutions to real world applications in the process industry.

Acknowledgment

References

ANSI/ISA–S95.00.01–2000, *Enterprise-Control System Integration Part 1: Models and Terminology*, ISBN: 1-55617-727-5, 2000, ISA–The Instrumentation, Systems, and Automation Society.

ANSI/ISA–95.00.02–2001, *Enterprise-Control System Integration Part 2: Object Model Attributes*, ISBN: 1-55617-773-9, 2001, ISA–The Instrumentation, Systems, and Automation Society.

ANSI/ISA-95.00.03-2005, *Enterprise-Control System Integration Part3: Activity Models of Manufacturing Operations Management*, ISBN: 1-55617-955-3, 2005, ISA–The Instrumentation, Systems, and Automation Society.

L.T. Biegler, I.E. Grossmann, 2004, *Retrospective on Optimization*, Comp. and Chem. Eng., 28, pp. 1169-1192.

T.F. Edgar, D.M. Himmelblau, L.S. Lasdon, 2001,*Optimisation of Chemical Processes*, McGraw-Hill Companies, Inc.: New York.

T. Fuchino, Y. Shimada, M. Miyazawa, 2008, *Business Process Model for Knowledge management in Plant Maintenance*, ESCAPE 18.

I.E. Grossmann, 2002, *Review of Nonlinear Mixed-Integer and Disjunctive Programming Techniques*, Optimization and Eng, vol 3.

P. Heijnen, Z. Lukszo, 2006, *Continuous improvement of batch wise operation—a decision support framework*, Production Planning & Control, 17 (4), pp. 355–366.

C.A. Méndez, *et al*., 2006, *State-of-the-art Review of Optimization Methods for Short-Term Scheduling of Batch Processes*, Computers and Chemical Engineering, vol 30.

M. Pinedo, 2001, *Scheduling: Theory, Algorithms, and Systems*, Prentice Hall.

B. Scholten, 2007, *The road to integration*, ISA, Reserarch Triangle Park.

D. E. Shobrys, D. C. White, 2002, *Planning, scheduling and control systems: why cannot they work together*, Computers and Chemical Engineering 26, pp. 149–160.

ISA Standard ISA-S88.01, 1995, *Batch Control. Models and Terminology*, ISA, Standard.

Integration Definition for Function Modelling (IDEF0), 1993, Publication 183, FIPS PUBS, 1993

M. Salverda, Z. Lukszo, P. Bosman, 2006, *A decision support tool for process optimization of sulphur free diesel production,* ESCAPE 16.

A.M. Klompenhouwer, Z. Lukszo, F. Janssen, 2007, *Analyzing the relationship between manufacturing lead-times and line flexibility – the Line Flexibility Model*, ESCAPE 17.

R.Ytsma, Z.Lukszo, R. Maliepaard, 2009, *Sustainable reduction of dredging fleet emission,* ESCAPE 19.

D. Stekelenburg, Z. Lukszo, J. Lowe, 2008*, Improvement of the production process of leached optical fibers in a technological and organizational context,* ESCAPE 18.

An Experimental Approach to Complement Process Systems Engineering Learning

Roger J. Zemp, Renata Waki and Flávio V. da Silva

School of Chemical Engineering
University of Campinas, UNICAMP
P.O. Box 6066, 13083-970, Campinas-SP, Brazil

Abstract

This paper describes a proposal for complementing Process Systems Engineering (PSE) undergraduate learning. Traditional teaching of PSE is normally restricted to modeling and computer simulations, with very little contact with actual processes. In this paper we describe how laboratory experiments can be combined with modeling and simulation tools to enhance the overall understanding of Process Systems Engineering.

Keywords: experimental process setup, modeling and simulation, undergraduate teaching

1. Introduction

Traditional teaching of PSE is typically focused on modeling and computer simulations, with very little contact with actual processes. We felt that applying theoretical concepts to practical problems would be very welcome by the students and improve their understanding of PSE. On the other hand, experimental work done by chemical engineering students in labs is more than often restricted to collecting data, doing some data processing and comparing the results to those expected by applying the appropriate theory.

We were looking for a way to combine experimental work with modeling and simulation tools. However, this should be done keeping low cost and low complexity in mind. This ruled-out experiments requiring measurement of composition, at least for this first tentative step. And since flow and temperature measurements are quite easy to be achieved, the heat exchanger experiment was chosen for our proposal.

2. The proposed approach

The heat exchanger experiment of our undergraduate lab suffered a major change a few years ago when the old double-pipe heat exchanger was scraped and a new small plate heat exchanger installed. Additionally, a hot-water storage heater was also installed to provide the hot fluid, while as cold fluid tap water was used. Flow measurements were done by variable area flow meters, and temperatures by thermocouples connected to a digital display. However, despite the changes in equipment the experimental work carried-out by the students was still restricted to the measurement of heat balance and heat transfer coefficient, and comparison with expected values. For our purpose this setup was far from ideal, and a few modifications were introduced.

2.1. The equipment

The new experimental setup consists of a process system (a heat exchanger) and a utility system (a cooling tower), which provides the cold water for the heat exchanger.

The heat exchanger is of the plate-type, with 19 plates and approx. 0.4 m² of area. Hot water (up to 65°C) is provided by a water storage heater.

The cooling tower is of the countercurrent type, made of three stacked PVC cylinders (40 cm diameter), using plastic mesh as filling in the middle cylinder. The bottom tank stores the cooled water, while the top cylinder contains the fan. Heated cooling water from the heat exchanger can be bypassed through a finned heat exchanger to pre-cool the water (as would be the case in an actual plant using air-coolers).

Pumps and valves are used to control the flow of water through the system.

2.2. Instrumentation

A set of industrial-standard instrumentation was used for measuring process variables: flows of hot and cold water (magnetic flow-sensors), fluid temperatures at inlet and outlet of exchanger, air-cooler and cooling tower (thermocouples), air humidity at inlet and outlet of cooling tower (humidity sensors), and pressure drop of cold water across the heat exchanger (differential pressure sensor). All instruments are connected to an industrial data-logging device using using 4-20 mA standards. Air flow through the tower was measured using a hot-wire anemometer.

Pictures of the actual equipments are shown in Figure 1.
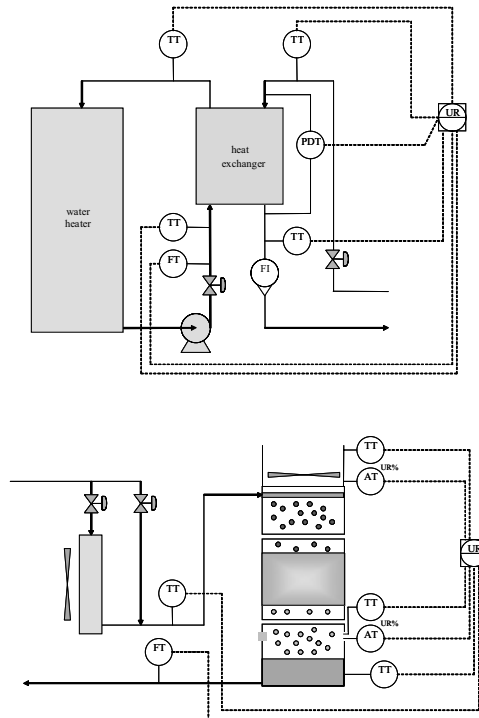
Figure 1 – The process (heat exchanger) and the utility system (cooling tower), with full instrumentation.

2.3. Data Monitoring

The measured process parameters are sent from the datalogger to a computer using the MODBUS/RS485 standard, and displayed to the user as process flowsheets. The flowsheets and additional screens were developed using the industrial-standard supervisory control and data acquisition system InduSoft Web Studio (Figure 2). Data can be stored in files for further off-line processing.
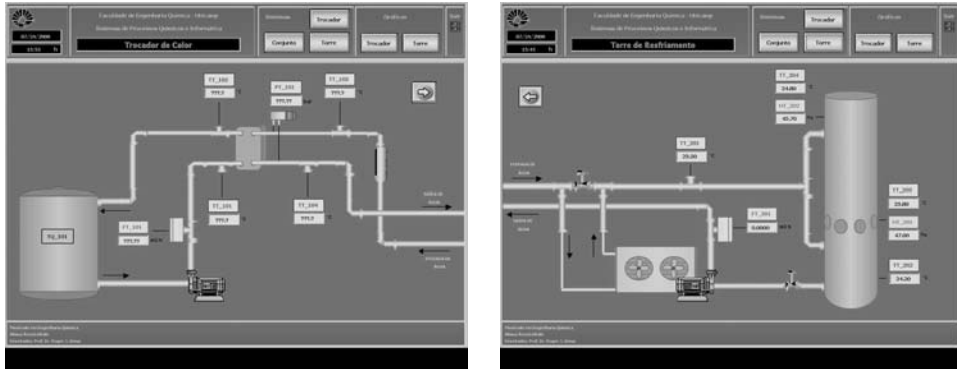
Figure 2 – Process flowsheets for the heat exchanger and the cooling tower

The overall benefits of the chosen experimental set-up are:
- The process is a small-scale example of a typical processing plant (a process and a utility system).
- Instrumentation and monitoring is done using industrial standard devices.
- Process parameter can be continuously monitored and stored.

3. Experiments

Once the experimental setup was finished, the practices to be used by the students were developed and implemented. As stated in the introduction, our goal was to combine experimental work with process systems engineering concepts, so we had to "set a scene" for each experiment. This was done by proposing an industrial problem to be solved using modeling and simulation techniques, but where some of the data required was not available and had to be determined experimentally. This was a major change in direction away from our traditional approach to lab teaching, where the experiment was the main issue. In our new approach, the problem to be solved was the main objective, and the lab providing one or more steps of the solution procedure.

Some of the proposed experiments are now presented, with the discussion of the relevant PSE techniques used.

3.1. Industrial heat transfer

The problem proposed to the students in one of designing a plate heat exchanger for a given process, with minimum cost. The process data (fluids, flows, temperatures) are known, as are the available pressure drops. The problem is that no correlation is available for estimating heat transfer coefficients, but they are to assume that the small-scale exchanger in the lab can be used to provide a suitable correlation.

As the first step the students have to implement a mathematical model of a plate heat exchanger, where for a given set of temperatures and flows the lab exchanger is simulated. This involves computing the energy balance and the actual overall heat transfer coefficient. Once their simulation is working, the students use the lab setup to perform a series of experimental runs to gather temperature, flow and pressure drop data. The next step is to propose a procedure to obtain the set of parameters that be fits the model to the experimental data. For this task a proper correlation for the heat transfer coefficient is chosen, normally of the form (Hewitt et al., 1994; Hewitt, G., 1993):

$$Nu = aRe^b Pr^c \qquad\qquad Eq.\ 1$$

Due to the restriction of only one fluid being available (water), it is suggested to use c=0.4. The students than have to implement their procedure to bet fit the proposed correlation to the experimental data:

$$\min \sum \left(U_{experimental} - U_{predicted}\right)^2 \qquad\qquad Eq.\ 2$$

A similar approach is used to obtain a correlation for the friction factor of the pressure drop equation. After a proper correlation for the heat transfer is obtained, the final step is to design a heat exchanger for the given problem. Here, different types of problems can be proposed: restrictions on exchanger area or number of plates, available pressure drop, required flow layout and number of passes. Normally the students have to resort to non-linear programming procedures to obtain a feasible result (Biegler et al., 1997; Edgar et al., 2001).

This type of task complements very well the concepts learned in the modeling, simulation and unit operation courses. Figure 3 shows a typical result from the correlation fitting procedure, with the points located close to the 45° diagonal.
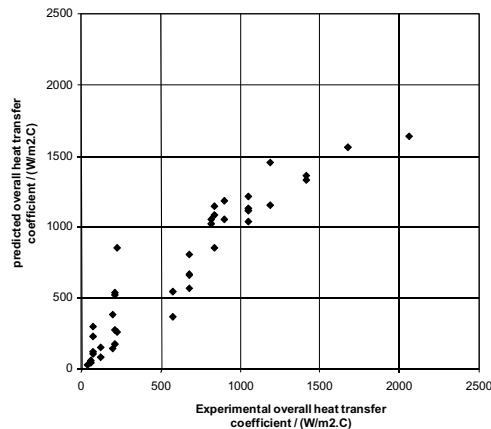
Figure 3 – comparison between measured and estimated global heat transfer coefficients

3.2. Process simulation and identification

In this problem the students are faced with the task to provide the Company's Engineering Department with a mathematical model of an existing heat exchanger. This model is supposed to be used to design a control system for a network of similar exchangers.

The actual work required is to build a mathematical model for the transient behavior of a countercurrent plate exchanger, and program it on a computer using an appropriate modeling system, MatLab for example (Silebi and Schiesser,1992). This requires the solution of distributed-parameter differential equations, and use of numerical methods. The students are then asked to gather experimental transient data in the lab, for several different process conditions, and compare these to the simulations results. Adjustments of the model are then done to match the model to actual data as closely as possible. The students have to deal with uncertainties in measured data, and be able to propose a method of fitting the transient model to the experimental data. Figure 4 shows a typical result from a simulation program, showing the behavior of the exchanger (exit temperature) for different process conditions and solution parameters.
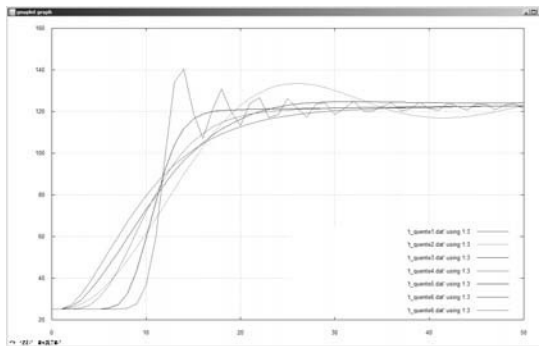
Figure 4 – transient behavior of heat exchanger

3.3. Interaction between process (heat exchanger) and utility system (cooling tower)

While the previous experiments dealt with a single unit operation, in this one the students are faced with the interaction between two unit operations: a process (heat exchanger) and a utility system (cooling tower).

The students are given the task to analyze the behavior of an industrial cooling tower under changes of process heat load. The details of the tower are all given, if an existing design, or need to be determined (if a new design). In both cases, however, the mass transfer coefficient is not known, and as in previous cases, a lab setup with similar tower filling is available for tests.

The students are initially required to implement a program that simulates the counter-current cooling tower. This requires the use of thermodynamics (psychrometry), and modeling techniques for countercurrent unit operations, with equilibrium and operating lines (Figure 5). The next step is to use the experimental setup to analyze the behavior of the cooling tower under different heat loads and cooling water return temperatures. The experimental data is the used to fit a very simple correlation for the overall mass transfer coefficient for the tower filling (Hensley, 2006).

In the second part of the experiment the students use the simulations and data obtained in the first part to solve the proposed industrial problem. Again, different problems can be proposed, with different process constraints or cost functions, so that each group of students has a different task to work on. In this experiment the students have to resort to modeling techniques involving differential equations / integration and optimization.
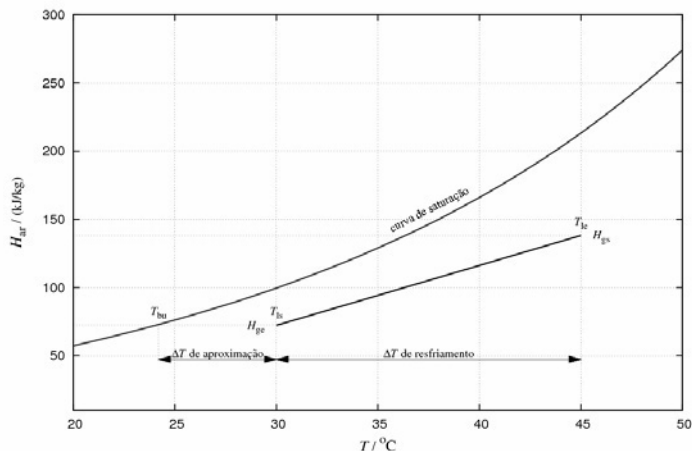
Figure 5 – Cooling tower operating line and saturation line

4. Conclusions

This paper describes the implementation of a new approach to combine experimental chemical engineering lab work with process systems engineering. Overall, this new experimental setup allows for the students to apply their knowledge of PSE in a more practice-oriented environment, dealing with actual process equipments. PSE tools like modeling and optimization are fitted into a problem-solving environment. Numerical tools like integration, differential equations and non-linear optimization are use to analyze and model actual processes. Instrumentation and data monitoring are done using industry-standard equipment.

Based on our assessment of the students that have been working under the new methodology, the proposed sequence of modeling a process, fitting the model parameters using experimental data, and using the model to optimize or design a given process has been shown to be very efficient in helping the students understand and apply process systems methodologies.

References

Biegler, L., Grossmann, I., Westerberg, A. 1997, *Systematic Methods of Chemical Process Design*. Upper Saddle River: Prentice Hall PTR.

Edgar, T., Himmelblau, D., Lasdon, L., 2001, *Optimization of Chemical Processes*. 2nd Edition, New York: McGraw-Hill Publishing Co.

Hensley, J. (editor), 2006, *Cooling Tower Fundamentals*, SPX Cooling Technologies, Inc., Overland Park, Kansas USA

Hewitt, G., 1993, Heat Exchanger of Design Handbook. S.L.: Begell House.

Hewitt, G., Shires, G., Bott, T. 1994, Process Heat Transfer. Boca Raton: CRC Press.

Silebi, C., Schiesser, W. ,1992, Dynamic Modeling of Transport Process Systems. Boston: Academic Press.

Cooperative Weblab: A Tool for Cooperative Learning in Chemical Engineering in a Global Environment

Galo A.C., Le Roux[a], Giann B. Reis[b], Charles D.F. de Jesus[b], Roberto C. Giordano[b], Antonio J.G. Cruz[b], Paulo F. Moreira Jr.[a], Claudio, A.O., Nascimento[a], Luiz V. Loureiro[b]

[a] *Chemical Engineering Department, Polytechnic School of the University of São Paulo, Av. Prof. Luciano Gualberto, trav. 3, 380, 05508-900, São Paulo, SP, Brazil*
[b] *Chemical Engineering Department, Federal University of São Carlos, Rodovia Washington Luis (SP 310), km 235 – 135 5 905, São Carlos, SP, Brazil*

Abstract

Weblabs are defined as a set of web-based physical laboratories that allows remote access and control in real time. Since the last decade, Weblabs are being more and more employed in many educational institutions around the world. In São Paulo state (Brazil) a "Cluster of Weblabs in Chemical and Biochemical Process Engineering" was implemented. This project has received a grant from FAPESP (State of São Paulo Research Agency). The experimental setups are physically placed in laboratories in São Paulo, São Carlos, Ribeirão Preto and Campinas, that are up to 250 km apart. This paper presents the implementation of two Cooperative Weblab (CW) experiments. The main concept behind the CW is to develop a set of experiments for undergraduate students that should be performed by "mixed teams" located in different institutions at the same time. Collaboration is achieved by gathering participants into working groups that are asked to simultaneously solve a technical problem, for which a weblab experiment is available. This procedure emulates challenges that will frequently take place in their future professional lives. Each group is supervised by a local instructor. Video conference software is used along the experiments (the institutions are connected through a high speed network, named KyaTera). All long the experiments, the students are encouraged to interact, exchange information and opinions on the phenomena that take place during the experimental practice. At the end of the session, the data collected are shared by the students and they produce a unique report for both groups. The use of this approach fosters interest in students. It is clear that the CW offer real benefits to chemical engineering education. It is now necessary to spread the dimension of the cluster and to increase the interaction with groups from other countries.

Keywords: Internet, experiment, teaching, intercultural

1. Introduction

In the year of 2004, the Foundation for the Support of Research in São Paulo State (FAPESP) established a program named Information Technology in the Development of Advanced Internet (TIDIA program). One of the projects of this program was named KyaTera, an acronym of Kya (net in "Tupi-Guarani", the language of the aboriginal people in Brazil) and Tera (from Terabits per seconds). KyaTera is a cooperative project

consisting in an optical fiber network designed for the research and development of high speed connections, linking various laboratories focusing on the study, development and demonstration technology and application on Advanced Internet. One of the projects of KyaTera is the "Cluster of Weblabs for Chemical and Biochemical Process Engineering". This project intends to build a network of undergraduate chemical engineering experiments that should be performed by "mixed teams", integrating students of four different Universities of the São Paulo state (EPUSP, Escola Politécnica da Universidade de São Paulo – the hub of the Weblabs network, UFSCar, Universidade Federal de São Carlos, USP-RP, Universidade de São Paulo, campus Ribeirão Preto and UNICAMP, Universidade de Campinas). Further information can be found at www.kyatera.fapesp.br. Figure 1 presents a picture of the network dark fibers implemented in the São Paulo state by KyaTera project.

Figure 1. Dark fiber connecting researchers laboratories in São Paulo state, Brazil.

Weblabs are defined as a set of web-based physical laboratories that permits remote and control access in real time (see Figure 2). During the experiments users do not need to be in front of the experimental apparatus to set-up experimental conditions. The changes are remotely made from computers with a graphical interface that permits access to the main variables of the process. At the end of the experiment, data can be transferred to the user.
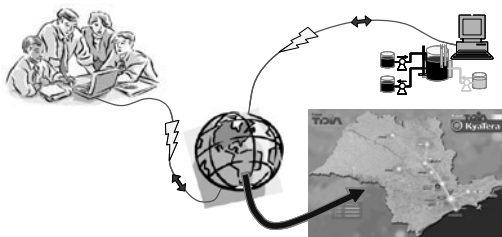
Figure 2. A sketch of the remote Weblab operation.

This paper presents the idea of implementing Cooperative Weblab (CW) experiments. The main concept behind the CW is to develop a set of experiments for undergraduate students that should be performed by "mixed teams" located in different institutions at the same time. The main idea behind this tool is to promote an increasing cooperative learning. The implementation of this tool encompasses the agreement of two or more institutions. A great effort is necessary in order to reach this kind of activity in the curricula of Chemical Engineering courses.

2. Review and Previous Works

The application of weblabs experiments in chemical engineering education is quite recent. The concept was introduced by Prof. Clark Colton at the heat transfer course of the undergraduate program at the Massachusetts Institute of Technology (MIT). The weblab experimental setup consists of a plate heat exchange with all necessary instrumentation and control devices. This weblab was shared with students from the chemical engineering from the University of Cambridge, UK, during their process control courses (Selmer et al. 2005). Another set of weblabs was built in University Leipzig by Ralf Moros under supervision of Helmut Rapp for training courses (personal communication).

SHIN et al (2002) were amidst the first authors to report a virtual web-based lab for unit operations course as a chemical engineering education tools. A collaborative web-based experiment was presented by GILLET et al. (2005) that provides more flexibility for students performing experiments in automatic control. They integrate learning environment that enables the recollection and sharing of preparatory notes and experimental results.

Two international workshops were organized one at MIT, in January 2005, and the other one at University of Cambridge, UK, in July 2005, with the participation of representatives from many institutions from various countries, e.g. Monterey Institute of Technology, Mexico, ENSIACET, France, University of Leipzig, Germany, discussed weblabs as an educational tool, its potential developments and outcomes.

3. Cooperative Weblab in Chemical Engineering

In this work we is introduce a unique approach for weblabs: Cooperative Weblab (CW). This new format for weblabs promotes intercultural experiences to students also allowing them to integrate the advantages of cooperative learning. Cooperative learning is a more formal kind of activity where students work in teams that stay together for extended periods of time under conditions that involve five criteria: positive interdependence, individual accountability, face-to-face interaction, development of

interpersonal skills, regular self-assessment of group functioning (Kaufman et al., 2000).

CW experiments must be performed by at least two groups of undergraduate students, in two different locations. The cooperation is achieved by asking the participants of the group to simultaneously solve a technical problem, for which a Weblab experiment setup is available.

4. Case Study: CW between USP and UFSCar

The CW presented here was selected among two weblabs of chemical engineering departments, one at USP and the other one at UFSCar, that are 255 km distant.

4.1. Mass Transfer Weblab at UFSCar

It is important for the education of chemical engineers, to handle the fundamentals of mass transfer involved in this experiment and to get familiar with techniques that assess rates of oxygen transfer from the gas phase into the liquid culture medium as well. It is known that the control of oxygen level during aerobic cultivation of microorganisms is very important, and this experiment is suited for this purpose. The oxygen can be removed from the system by bubbling nitrogen, or added by bubbling oxygen. The dissolved oxygen is monitored by a specific electrode. After reaching zero oxygen concentration, the nitrogen flow stops and air starts flowing. The electrode probe measures the dissolved oxygen (DO) in the liquid phase. The experiment aim is to calculate kLa values at different operating conditions of air flow rate and stirrer speed employing the gassing-out method (Shuler and Kargi, 2002; Blanch and Clark, 1997). The main screen of the weblab for mass transfer was built employing National Instruments hardware for data acquisition and LabVIEW software as the supervisory system. The users can choose the experiment operating conditions (air flow rate and stirrer speed) through the main screen. The users can choose the experiment operating conditions either (air flow rate and stirrer speed) through this interface.

4.2. Reactor Temperature Control Weblab at USP

The reactor is a continuous stirred tank reactor (CSTR) and was assembled specially for teaching purposes. The CSTR is a 4.5 L capacity, jacketed glass vessel. At the inner part of the reactor there is an immersion heater (2.500 W), a stainless steel coil, a temperature sensor, a level sensor, a pH sensor, and a double helix impeller. The reactor temperature can be controlled by manipulating the hot water flow through the coil or the cold water flow through the jacket by two electro pneumatic valves. An electrical heater, immersed in the reactor, can simulate different exothermic reaction behaviors. The interface is very friendly and includes a webcam in order to remember the users that they are dealing with an actual setup and not with a simulator. During the CW experiment, only one user at a time can change the system settings, monitored by the other users. The users can also communicate with each other with video conference facilities, and, if necessary, they download the data log and evaluate the results during or after the experiment.

During the experiments, the students must perform calculations together in order to apply the Cohen-Coon (or other equivalent) procedure (Seborg et al. 1989) to obtain the PID tuning parameters. The set of PID tuning parameters is implemented by closing the control loop and is tested by different methods, e.g., set point changes and disturbance rejection tests.

5. Performing the CW Experiments

The main goal of CW experiments is to make students from the different institutions work together. Groups are formed with students from UFSCar and USP in equal number. The idea is to create a link between these students. Firstly they are forced to work together in order to prepare the experiment. They must review the theory and talk in order to decide what procedure will be used during experiments. The idea is not to give them just a recipe but to only give them the guidelines, so that they are forced to decide what will be the sequence of the experiment by themselves. If necessary a previous video conference is performed between the students and at least one instructor.

The experimental session is performed with a lot of video streaming and conversation. There is one instructor at each side, and the main function is to serve as a consultant and to assist them with practical issues that could arise during the experiment. The students exchange information and opinions about the phenomena that take place. In the mass transfer experiment, some concerns arise about the quality of the mixing, the size of the bubbles and the various problems that arise during the experiment (bubbles blocking the tip of the probe, high coalescence phenomena, and conditions of inefficient mixing). In the reactor temperature control experiment, a step change on the coil flow is performed in order to obtain the response curve, and the students are frequently concerned about if the system has reached the new steady state. In general they are surprised by the great time constants of the system.

After the experiments, the students are invited to freely interact with the equipment. They are left free to explore extreme conditions and to test different perturbations. In the case of the reactor temperature control the student can set different unmeasured perturbations by means of a resistance that is immerse in the reactor and verify if the controller they set up is able to cope with them.

The interaction that takes place between the students during the experiment is very strong. As the students shall process the data and prepare the final group report together, the interaction is still present and they are invited to use the network to communicate and prepare the report.

A closing video conference session is performed where the students get a feedback from the instructors and discuss the main issues of the experiments.

6. Conclusions

CW experiments, as any other cooperative learning experiment is sought to allow students the development of important skills, as team working and a significant learning in some relevant chemical engineering topics.

The opinion of the students is in general extremely positive and encouraging. The aforementioned cluster already performed CW experiments with other institutions like ENSIACET in Toulouse, France and TH Karlsruhe, Germany. The experience has shown that this is a valuable educational tool in a world with an increasing global interaction between professionals.

A special effort is necessary in order to fit this kind of activity in the curricula of chemical engineering programs, because it involves two or more institutions. This is the reason why it is fundamental to make this experience known to a broad public: to congregate a great diversity of centers that could be open to perform this kind of experiment.

Some practical issues must be overcome for CWs in different countries such as the difference in time zones, calendars and language. Nevertheless these issues are

unimportant in comparison with the worst of the intercultural experience CW can bring to the chemical engineering education.

References

Blanch, H. W., Clark, D. S., Biochemical Engineering. Marcel Dekker, Inc., New ork, 1997.

Gillet, D., Ngoc, A.V., Rekik, ., Collaborative Web-Based Experimentation in Flexible Engineering Education, IEEE Transactions on Education, 48(4), November 2005.

Kaufman, D.B.,Felder, R.M.,Fuller, H., Accounting for individual effort in cooperative learning teams, Journal of Engineering Education, 89 (2), 33-140, 2000

Kyatera Program Annual Reports, www.kyatera.fapesp.br

Seborg, D.A., Edgar, T.H., Mellichamp, D.A., Process dynamics and control, John Willey Sons, New ork, 1989.

Selmer, A., Goodson, M., Kraft, M., Sen, S., Faye Mcneill, V., Johnston, B.S., Colton, C.K., Performing Process Control Experiments Across the Atlantic, Chemical Engineering Education, 39 (3), Summer 2005.

Shin, D., oon, E.S., Lee, K. ., Lee, E.S., A web-based, interactive virtual laboratory system for unit operations and process system engineering education: issues, design and implementation, Computer and Chemical Engineering, 26, 319-330, 2002.

Shuler, M. L., Kargi, F., Bioprocess Engineering: Basic Concepts. 2nd edition, Prentice Hall International Series, Upper Saddle River, NJ, 2002.

Acknowledgement

Steps for a Multidisciplinary Engineering Competition

Fernanda E. S. Duarte[a], Caio V. P. Delgaudio[a], Flavia N. David[a], Flavio Waltz[a] and Valter C. Souza[a]

[a]*CHEMTECH – A Siemens Company, Rua da Quitanda, n. 50, Rio de Janeiro, CEP 20011-030 - Brasil.*

Abstract

The worldwide number of engineering projects is growing as never before and so is the demand for trained workers. These facts motivate CHEMTECH to train engineering students in crucial areas from process industry, contributing to their development, improving their acceptance in the labor market, and guarantying the availability of better professionals to the industry. In this context, CHEMTECH promoted a multidisciplinary competition between engineering undergraduate students from the most prestigious universities of Brazil. In 2008 the National Engineering Marathon included Computational Fluid Dynamics and Process Control Theory, two important engineering areas. The application of CFD (Computational Fluid Dynamics) to engineering has been growing extremely fast during the last years, especially in design and process optimization as well as to prevent and to mitigate potential process risks. CFD greatest appeal is that one makes these kinds of studies without interrupting the normal operation of industrial plants. Process control analysis is as important as CFD because an efficient control system guarantees stable and confident processes, improving the quality of the final product. This challenge was organized in three steps: call for the contest, training of the attendees and the contest itself, which took place during the Rio Oil Gas Expo Conference. During the first step, CHEMTECH invited the best Brazilian universities, each one represented by their best students and an advisor professor. In the next step they were all trained through the internet using the methodology of Distance Learning. The students were trained in PHOENICS (Parabolic Hyperbolic or Elliptic Numerical Integration Code Series), a CFD software developed by the British company CHAM (Concentration, Heat and Momentum Limited) and Scilab, open software for Control Process. This training also gives the students an opportunity to be in contact with real industrial applications. In the end of this step two students from each university, as a team, were qualified for the final contest, in which they were presented to real industrial problems that they needed to solve using the learned tools. Their solutions were evaluated by professionals from CHEMTECH and from some of its remarkable clients and the best group was the champion. Beyond the knowledge and recognition, the students and professors of the winning group also receive prizes and their university receives a donation of computers. Moreover, CHEMTECH is contributing to improve the national engineering and the quality of projects in world process industry.

Keywords: Multidisciplinary Engineering Competition; Computational Fluid Dynamic; Process Control Theory; and Phoenics.

1. Introduction

The worldwide number of engineering projects is growing as never before and so is the demand for trained workers. These facts motivate CHEMTECH to train engineering students in crucial areas from process industry. This multidisciplinary competition improves their acceptance in the labor market, as training the students is the main objective of this project.

CHEMTECH has been promoting and sponsoring events involving Brazilian university students since 2004. This kind of competition develops the technical knowledge of the students and helps them to apply this knowledge in real problems and common industrial situations.

The first edition was promoted in 2004. A regional competition was conduced as a first step, in three Brazilian cities: Belo Horizonte (Minas Gerais state group), Rio de Janeiro (Rio de Janeiro state group), and São Paulo (São Paulo state group). The competition was a really success, fact that encouraged the organization of a national competition.

In 2006, Chemtech organized the first National CFD challenge. The event counted with the participation of thirteen Brazilian universities and more than thirty students. The competition took place in the Rio Oil   Gas Expo of that year. This well-succeeded experience was expanded for other areas of the knowledge, as process control analysis. The first process control challenge happened in 2007. In 2008, a innovation was proposed: Organize a challenge mixing both areas, the National Engineering Marathon, involving eighteen universities and more than two hundred students.

2. Competition

The National Engineering Marathon included Computational Fluid Dynamics and Process Control Theory, two important engineering areas. The application of CFD (Computational Fluid Dynamics) to engineering has been growing (due to advancements in computer's process capacity), especially in design and process optimization as well as to prevent and to mitigate potential process risks. CFD greatest appeal is that one makes these kinds of studies without interrupting the normal operation of industrial plants. Process control analysis is as important as CFD because an efficient control system guarantees stable and confident processes, improving the quality of the final product. This multidisciplinary competition contributes to student's development, guarantying the availability of better professionals to the industry. This challenge was organized in three steps: call for the contest, training of the attendees and the contest itself, which took place during the Rio Oil   Gas Expo   Conference 2008.

2.1. Call for the contest

During the first step, CHEMTECH invited eighteen Brazilian universities: UFRGS - Federal University of Rio Grande do Sul, UFSC - Federal University of Santa Cartarina, USP- São Paulo University, Unicamp - University of Campinas, UFSCar- Federal University of São Carlos, ITA - Technological Institute of Aeronautical, UFBA - Federal University of Bahia, UFMG - Federal University of Minas Gerais, UFU - Federal University of Uberlandia, UFCG - Federal University of Campina Grande, UFPE - Federal University of Pernambuco, UFRN - Federal University of Rio Grande do Norte, UFPR - Federal University of Paraná, UFRJ - Federal University of Rio de Janeiro, UFF - Fluminense Federal University, IME - Engineering Military Institute,

UERJ - State University of Rio de Janeiro, UFES - Federal University of Espírito Santo, each one represented by an advisor professor and their best students, which means 250 graduation students. The students were from different engineering areas. The distribution is showed in the Figure 1.

Engineering Areas Distribution
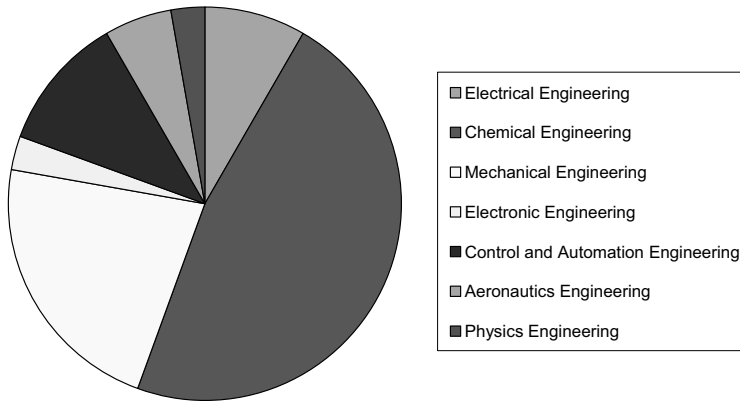
Figure 1: Distribution of the students trough engineering areas

2.2. Training

In the next step, the students were all trained through the internet using the methodology of Distance Learning. The course material was composed by electronic books, divided in modules, available for download to all students. Each module corresponded to a week of course. During this time, chats among the students and the monitors (CHEMTECH professionals, with experience in the CFD and Control fields) were scheduled. Moreover, the student's doubts and questions could be discussed through forums on the internet or trough email, directed to the monitors.

The most difficulty task in this step, and in this teaching methodology itself, is to keep the attendance and participation levels. In order to try to keep the levels, the proposed exercises were all based in real industrial problems and special chat sessions with remarkable specialists and professors in those areas were scheduled. Although all of this efforts the participation level lowered along the course. In the first week, the attendance was 92% (average over the eighteen universities), while in the last week it was 48%. The tendency described above can be observed in Figure 2, which shows the participation along the course weeks for four selected universities.
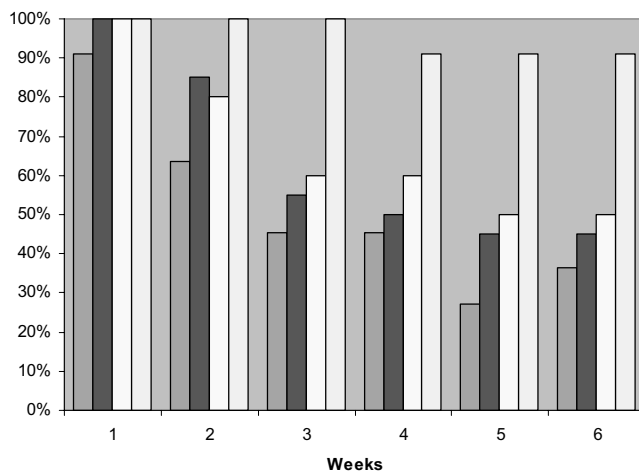
Figure 2: Participation in the training along the weeks. Each color indicates a different university and the participation level is indicated through the bars.

The whole training took six weeks. Besides the course material available about CFD and Control theory, two software were sent to the students: PHOENICS (Parabolic Hyperbolic or Elliptic Numerical Integration Code Series), CFD software developed by the British company CHAM (Concentration, Heat and Momentum Limited) and Scilab, open software for Control Process. In order to make easy the learning process with the software, some tutorials were written and sent to the students, concerning the main functions of the software.

An exercise list was also sent every week to the students, with problems involving the topics learned until that point. The solution of the list was used to rank the students of each university in order to select the better grades to the next step of the event: the contest itself. Doubts and questions about the exercises could also be clarified during the weekly chats with CHEMTECH's professionals.

This training gave the students an opportunity to be in contact with industrial applications, something that occurs rarely during the under graduation courses. In the end of this step, two students from each university, as a team, were qualified for the final contest, during which the students were presented to real industrial problems that they needed to solve using the learned tools.

2.3. Final Contest

The final contest took place in the Rio Oil   Gas 2008, the biggest event of the Oil and Gas industry in Latin America. The selected students and their professors were brought to Rio de Janeiro from many places in Brazil during the exposition week.

This final step was divided in two parts. The first one was constituted by eliminatory tests, which were done by all the eighteen teams. The six better teams, ranked by the grade in the tests, were classified to the second step. A new chance was given to the twelve teams that did not get their place in the second step: another test was applied and the two best teams were also classified. With the best sixteen from the initial two

hundred and fifty students the second step was conducted in the last day of the Rio Oil Gas. The final test lasted four hours and a problem about a solar energy collector and its functioning coupling CFD and control was proposed to the students.

Their solutions were evaluated by professionals from CHEMTECH and from some of its remarkable clients and the best group was the champion. Beyond the knowledge and recognition, the students and professors of the winning group also received prizes and their university received a donation of computers.

3. Conclusion

The National Engineering Marathon promoted an exchange among universities, their students, and the engineer companies, identifying new talents. Moreover, CHEMTECH is contributing to improve the national engineering and quality of projects in world process industry.

Some recognition has already resulted from this educational initiative and the company was awarded the 2004 Social Responsibility Prize, granted by CREA (National Engineering and Architecture Council).

The competition has also awaked the interest of a lot of undergraduate students in the most different engineering areas and has revealed some great potential future engineers. This successful experience is now being expanded to other countries, in the International Engineering Marathon that will happen in this year.

This page intentionally left blank

A Flexible Laboratory-Scale Quadruple-Tank Coupled System for Control Education and Research Purposes

João B. M. Santos[1], George Acioli Júnior[1], Henrique C. Barroso[1] and Péricles R. Barros[1]

[1]*LIEC – Department of Electrical Engineering, CEEI, UFCG*
Av. Aprígio Veloso, 882 Bairro Universitário CEP 58429-900, Campina Grande – PB, Brazil

Abstract

In this paper it is described a flexible laboratory-scale quadruple-tank coupled system designed for Control Education and Research purposes. The quadruple-tank process presented is a flexible structure for experimental implementation and evaluation of different control strategies in a multivariable non-linear process. The structure allows the implementation of different experiments, such as SISO (single-input single-output), TITO (two-input two-output) and MIMO (multiple-input multiple-output), with the option of coupling or cascade effect between tanks. A general description for the system is presented. A linear dynamic model of the tank system is presented for a SISO configuration and simple experiments are shown which illustrate the use of the laboratory-scale system in undergraduate control course.

Keywords: Control Education, Laboratory Plant, Coupled Tanks, PID Control, System Identification, Dynamical Modeling

1. Introduction

Laboratory-scale processes are important tools that help in the academic study of various phenomena observed in industrial plants, especially in the study of techniques for the multivariable control of non-linear processes.

The control of the liquid levels in storage tanks and reaction vessels is a common control problem in chemical and petrochemical processes industries. The tank process can be modeled with some constraints, allowing a quick and efficient implementation and test of different strategies of control.

Some multivariable laboratory-scale tanks systems are commercially available, for example from uanser Consulting in Canada, Educational Control Products in the United States, and Feedback Instruments and Tec uipment in the United Kingdom (Johanson, 2000).

In this paper, a new flexible laboratory-scale quadruple-tank system for experimental level and flow control is described. This system has characteristics of construction and application rather individual, becoming a very flexible platform for implementation and evaluation of different control strategies in multivariable non-linear processes. The laboratory setup is used in Control and Automation undergraduate courses at the DEE UFCG (Departamento de Engenharia Elétrica Universidade Federal de Campina Grande) and for control research purposes. A laboratory-scale quadruple-tank system photo is shown in Figure 1(a).
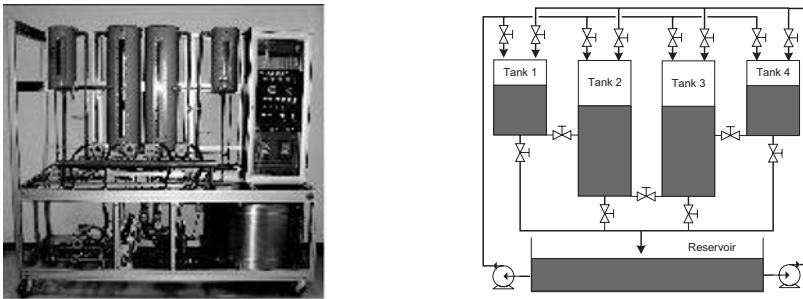
Figure 1: (a) Laboratory-scale Tank System (b) schematic diagram of the tank

The laboratory-scale quadruple-tank system gives the students the opportunity to solve control problems as well as to integrate theoretical knowledge obtained at lectures with practical experience. The experiments performed in the laboratory system for teaching undergraduate students in modeling, continuous-time identification and PI PID controller design are presented here. For illustration, in this paper only experiments performed in SISO (Single-Input Single Output) configuration are presented.

This paper is organized as follows. In section 2, the laboratory-scale quadruple-tank system is described. In section 3, a linear dynamic model of the SISO tank system presented in literature is reviewed. In section 4, some of the experiments performed in the laboratory system are described. In section 5, the experiments results are presented and, finally, conclusions are presented in Section 6.

2. Laboratory-Scale System Description

The laboratory-scale system (shown in Figure 1(a)) consists of four tanks of different sizes, two hydraulic pumps, two frequencies inverters, six differential pressure transmitters, a PLC and a PC (Personal Computer) with SCADA (Supervisory Control and Data Acquisition).

Four differential pressure transmitters are used for measuring the level of the liquid in each tank. Two are used for measuring the liquid-flow of the hydraulic pumps. The primary element for flow measurement is the orifice plate. The two hydraulic pumps are driven by frequency inverters. This way, the liquid is pumped into the top of each tank. A simplified schematic diagram of the quadruple-tank system is shown in Figure 1(b).

The laboratory-scale system allows multidisciplinary use to support several disciplines at the electrical engineering undergraduate course at UFCG. The undergraduate students have the opportunity to test modeling and simulation techniques, linear control systems design, PLC (Programmable Logic Controller) programming, supervisory design and industrial instrumentation. It is also used in the graduate course for experimental design, validation of SISO and TITO control techniques, system identification and linear modeling of non-linear process.

Solenoid valves are installed to ensure the physical integrity of the plant when an abnormal condition is detected. They interrupt the liquid-flow from the pumps to the tanks. An alternative way had to be created because the pumps are not turned off.

The control is implemented in an Allen-Bradley's Controllogix PLC or using Matlab with its OPC (OLE for Process Control) toolbox. To improve the interface between the user and the platform for the PLC, a supervisory window turns available using Wonderware Intouch SCADA software. The entire configuration parameters,

monitoring and any change in the process control is done using the supervisory. The PLC turns available the process data for the supervisory through the RSlinx OPC Server.

The arrangement of the set of valves allows the configuration of the plant for carrying out different experiments, such as SISO, TITO and MIMO, with the option of coupling or cascade effect between tanks.

In this paper a SISO configuration schema is adopted. In this schema, the liquid is pumped into the top of the second tank from the reservoir, and outflows of the tank only due to gravity. The general objective of the control is to reach and stabilize the level in the tank 2 by controlling the entry flow.

3. Dynamic Model

In this section, the dynamic model for the SISO configuration system is presented (Ogata, 2003). Industrial processes generally have turbulent flux resulting in non-linear model equations. So, linearization is used to simplify the analysis and controller design. The resistance R to the liquid flux is defined as the level variation ΔH necessary to cause a unit variation on the outflow ΔV. The capacitance C of a tank is defined as the liquid quantity variation necessary to cause a unit variation in the liquid level.

Assuming the outflow turbulent, the steady state flow is $Q = K\sqrt{H}$ where, Q is the flow in m^3/s, K is the coefficient in $m^{2.5}/s$ and H is the level in m.

In the SISO configuration q_i is the entry flow in m^3/s, q_o is the outflow in m^3/s, V_I is the entry valve, V_o is the out valve and C_I is the capacitance. Small letters are used for the small deviations in the stead state values of the correspondent variables.

We have

$$C_1 \frac{dh}{dt} = q_i - q_o \qquad (2)$$

and

$$q_0 = 2\frac{h}{R_o} \qquad (3)$$

where, R_o is the out valve resistance. Substituting Eq. (3) in Eq. (2) results

$$\dot{h}(t) = -\frac{2h}{R_o C_1} + \frac{1}{C_1} q_i \qquad (4)$$

Eq. (4) is the system dynamic equation. This equation relates the liquid level variation with the incoming input.

4. Experiments Description

4.1. Standard Relay Experiment – Gain Margin Estimate

The standard relay test presented in (Åström,1995) is used to estimate the critical point gain and frequency. It can be shown (Schei, 1994) that if this relay test is applied to a closed loop system, with transfer function $G(s)$, the limit cycle occurs at the closed loop critical frequency and the gain margin can be computed from the loop gain

$$L(j\omega_{gm}) = G(j\omega_{gm})C(j\omega_{gm}) \qquad (8)$$

4.2. Loop-Gain Relay Experiment - Phase Margin Estimate

A general relay procedure to estimate the frequency point for which a given transfer function has a desired gain is presented in (Arruda, 2003a). The conditions of the limit cycle operation are defined by the following proposition, with proof found in (Arruda, 2003a):

Proposition 1: Consider a stable closed loop *M(s)*, with loop-gain *L(s)*, and a real positive number *r*. So the transfer function

$$F(s) = \frac{2}{r} \frac{M(s)}{M(s)\left(\frac{1-r}{r}\right)+1} - 1 \qquad (9)$$

is also stable. Then if a limit cycle is present, it oscillates at a frequency ω_o such that

$$|L(j\omega_0)| \approx r \qquad (10) \quad \blacksquare$$

4.3. Relay Based Gain and Phase Margins Redesign

The closed-loop performances are evaluated here on the phase and gain margins sense using the relay estimations previously described. This information is used to redesign the controllers as proposed in (Arruda, 2003b). This problem is solved using an iterative approach applied to the following equations:

$$\angle G(j\omega_u)C(j\omega_u) = -\pi$$

$$|G(j\omega_u)C(j\omega_u)| = \frac{1}{A_m} \qquad (11)$$

$$|G(j\omega_g)C(j\omega_g)| = 1$$

$$\angle G(j\omega_g)C(j\omega_g) = -\pi + \phi_m$$

where A_m is the desired gain margin and ϕ_m the desired phase margin. The iterative algorithm only requires the knowledge of the frequencies ω_u and ω_g. These frequencies are the solutions to Eq. (11a) and Eq. (11c). This algorithm uses the following lemmas to update the controller's parameters.

4.3.1. Controller Redesign for Gain Margin

The controller gain can be calculated for achieving the gain margin A_m using Eq. (11b). That is, with the current gain margin GM^k and the critical frequency ω_u^k one can compute the controller proportional gain $\overline{K}_c^{k+1}$ from

$$\overline{K}_c^{k+1} = \frac{K_c^k GM^k}{A_m} \qquad (12)$$

4.3.2. Controller Redesign for Phase Margin:

The controller gain can be calculated for achieving the phase margin ϕ_m using Eq. (11d). This step is separated into two parts:

1. Determine T_i^{k+1} such that Eq. (11d) is satisfied, i.e.,

$$T_i^{k+1} = \frac{\tan -\pi + \phi_m - PM^k + \tan^{-1}(\omega_g^k T_i^k)}{\omega_g} \qquad (13)$$

2. Now, update the controller proportional gain K_c^{k+1} such that the loop gain at the

   frequency ω_g^k is equal to one,

$$K_c^{k+1} = \overline{K}_c^{k+1} \frac{\sqrt{(1 \ T_i^k)^2 + \omega_g^2}}{\sqrt{(1 \ T_i^k)^2 + \omega_g^2}} \qquad (15)$$

The controller at the end of the iteration is finally given by

$$C^{k+1} = K_c^{k+1} \left(\frac{s+1 \ T_i^{k+1}}{s} \right) \qquad (16)$$

5. Experimental Results

5.1. Controller Desing

There exist several design techniques based on the FOPDT model (Åström, 1995). Consider the Chien, Hrones and Reswick (CHR) SetPoint 0% and 20% methods. In these techniques the PI controller is designed based on the parameters τ, T_p, K_p and $\theta = \tau/T_p$. Two PI controllers are designed.

$$C_{chrset0\,pi}(s) = 182.45 \left(1 + \frac{1}{1213.92s} \right)$$

$$C_{chrset20\,pi}(s) = 312.77 \left(1 + \frac{1}{1011.6s} \right)$$

The closed-loop step responses are shown in Figures 6(a) and 6(b). The closfed-loop with $C_{chrset0\,pi}(s)$ is more aggressive than closed-loop with $C_{chrset20\,pi}(s)$.

5.2. Closed-Loop Performance Evaluation

The closed-loop performance is evaluated here on the phase (PM) and gain (GM) margins sense using the standard relay and loop gain experiments. The margins estimated for the two closed-loops are in Table 1. The relay experiments in closed-loop with $C_{chrset20\,pi}(s)$ are shown in Figures 8(a) and (b).

	GM	PM
$C_{chrset0\,pi}(s)$	12.16	12.68
$C_{chrset20\,pi}(s)$	8.96	25.05

Table 1: Margins for the two controllers

6. Conclusion

In control engineering courses, a laboratory experience that complements theoretical lecture material is essential for students. In this way, a flexible laboratory-scale quadruple-tank system and a set of experiments have been presented. The laboratory system is used in Control and Automation courses and in control researches at DEE UFCG. In this paper only simple experiments performed in undergraduate courses are presented.
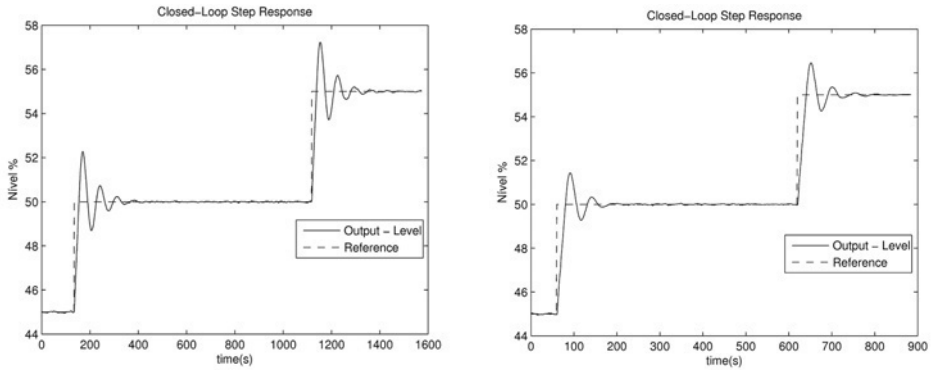
Figure 2: (a) CHR controller closed loop step response (set point 0%) and (b) CHR controller closed loop step response (set point 20%)
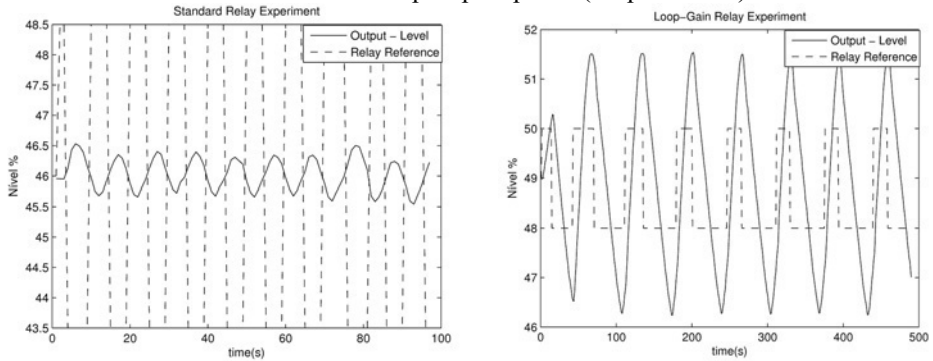
Figure 8: (a) standard relay experiment and (b) loop-gain relay experiment

References

Åström, K. J. and HÄagglund, T. (1995). PID Controllers: Theory, Design and Tuning, 2nd edn, Instrument Society of America, Re-search Triangle Park, North Carolina.

Coelho, F. S. and Barros, P. R. (2003). Continuous-time identification of first-order plus dead-time models from step response in closed loop, 13th IFAC Symposium on System Identification, Rotterdam (The Netherlands).

de Arruda, G. H. M. and Barros, P. R. (2003a). Relay-based closed loop transfer function frequency points estimation, Automatica 39(2): 309-315.

de Arruda, G. H. M. and Barros, P. R. (2003b). Relay based gain and phase margins PI controller design, IEEE Transctions on Inst. And Meas. Tech. 52(5): 1548-1553.

Johansson, K. H. (2000). The quadruple-tank process: A multivariable laboratory process with an adjustable zero, IEEE Transactions on Control Systems Technology 8: 456-465.

Ogata, K. (2003). Engenharia de Controle Moderno, Prentice-Hall, Inc.

Schei, T. S. (1994). Automatic tuning of PID controllers based on transfer function estimation, Automatica 10: 1983-1989.