Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Estimation of Distillation Compositions Using Sensitivity Matrix Analysis and Kernel Ridge Regression

Qi Li, Cheng Shao

Institute of Advanced Control Technology, Dalian University of Technology, Dalian Liaoning 116024, China, (e-mail: liqidut@gmail.com, cshao@dlut.edu.cn).

Abstract: The stringent quality requirement of petroleum products in highly competitive markets makes on-line controlling of distillation composition essential. In this paper, a novel method using sensitivity matrix analysis and kernel ridge regression to implement on-line estimation of distillation compositions is proposed. In the approach, the sensitivity matrix analysis is presented to select the most suitable secondary variables to be used as the estimator's inputs. The kernel ridge regression is used to build the composition estimator. The influence of measurement noise on the estimator's performance is also investigated. Application to a simulated distillation column demonstrates the effectiveness of the method.

1. INTRODUCTION

The stringent quality requirement of petroleum products in highly competitive markets makes on-line controlling of distillation composition essential. But unfortunately few hardware sensors are available on-line to distillation compositions. Many hardware sensors such as gas chromatographs and NIR (Near-InfraRed) usually possess significant time lags and high investment and maintenance costs (Kano, 2000, 2003). The best way to solve this problem is building composition estimators or soft sensors (Joseph and Brosilow, 1978; Mejdell and Skogestad, 1991a, b; Kano, 2000, 2003; Zamprogna, Barolo, and Seborg, 2004, 2005).

There are two methods to build the composition estimators. One way is to build the mechanism model on-line. However, it is often difficult in refineries, due to the complexity of industrial distillation processes. Physical modelling can be very time-consuming and significant parameters are generally unknown. The other way is to adopt empirical model. There are many algorithms have been used to build the estimator on this way, including using multivariate regression analysis (Mejdell and Skogestad, 1991a,b; Jie Zhang, 2001; Kano, 2000, 2003), artificial neural networks (Bhartiya and Whiteley, 2001), support vector machine regression (Yan and Shao, 2004) etc.

Due to the strong correlation among tray temperature measurements of the distillation column, composition estimators based on principal component analysis (PCA) and partial least squares (PLS) regression have been widely used. However, large samples are needed in these methods, and models are insensitive to measurement errors. Moreover, due to the nonlinear of composition estimators, many composition estimators based on artificial neural networks have been proposed and successfully applied in industrial processes. However, there are no guarantees of avoidance of local minima, the overfitting phenomenon and the number of hidden units in general neural networks are usually difficult to choose (Yan and Shao, 2004).

Optimal selection of the secondary variables is very important to the composition estimator. Joseph and Brosilow (1978) suggested an iterative selection method based on the addition of temperature measurements to the optimal set, one at a time. However, when the number of secondary variables is large, this method may be time consuming. In order to optimally select the secondary variables for the composition estimator, Zamprogna, Barolo, and Seborg (2005) present principal component analysis to choose the secondary variables in a batch distillation column. They defined a sensitivity matrix to measure the sensitivity of temperatures and the batch time is discussed.

In this paper, we focus on the algorithm of optimal selection of the secondary variables and building the composition estimator for a continuous distillation column. We take a novel method base on sensitivity matrix analysis and kernel ridge regression (KRR) to build the composition estimators. Ridge regression (RR) is a classical statistical algorithm which have been known for a long time (Hoerl and Kennard, 1970a,b). However, when RR deals with the nonlinear regression, it will encounter the "curse of dimensionality" problem (Saunders, 1998). We describe a nonlinear version of the RR, which allows the use of kernel functions. Compared to the traditional RBF neural network, using the same sample data, the simulation results show that the composition estimator based on kernel ridge regression has better abilities of model generalization.

The paper is organized as follows. Section 2 presents the distillation process model and the conditions of dynamic simulation. The sample data for the composition estimator is collected. Section 3 states the sensitive matrix analysis algorithm and section 4 provides details on the novel kernel ridge regression algorithm. Section 5 presents the optimal selection of the secondary variables for the composition estimator. Then the most sensitive variables are chose to build the composition estimator by using the KRR algorithm. Moreover, we discuss the effect of measurement noise on the

estimator's performance. Finally, conclusions are given in Section 6.

2. THE DISTILLATION PROCESS

2.1 First-principles dynamic model of distillation column

In this paper, a continuous distillation column is chosen as the research object which is well known for being used in composition estimator performance studies (Mejdell and Skogestad, 1991a). The schematic diagram of the continuous distillation column is shown in Fig. 1.

The column consists of 41 theoretical trays including a total condenser and a reboiler. A first-principles dynamic model is considered as the representation of the continuous distillation column. Unlike other models used for similar studies (Quintero, Luyben, and Georgakis, 1991), this model considers varying molar holdups on each tray, that is, it includes liquid flow dynamics. The internal liquid rate L_m on stage m is determined by means of the linearized relationship:

$$L_i = L_{0i} + (M_i - M_{0i})/\tau_L + (V_{i-1} - V_{0,i-1})\lambda \qquad (1)$$

Where L_{0i} [Kmol/min] and M_{0i} [Kmol] are the nominal values for the liquid flow and holdups on stage i. This means that it takes some time from we change the liquid in the top of the column (L_T) until the liquid flow into the reboiler (L_B) changes. The energy balances are not included in the dynamic model; therefore, the vapor rate is constant ($V_{i-1} = V_{0,i-1}$) inside the column. Other assumptions are ideal trays, well-mixed capacities, boiling feed, total condensation with no subcooling, negligible heat losses, and constant pressure operation. The base stead-state condition is summarized in Table 1.

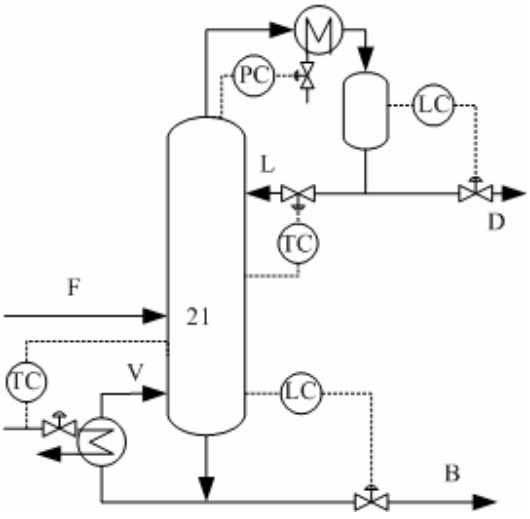

Fig 1. Schematic control scheme of distillation column

Table 1. Stead-state conditions of distillation column

No. of theoretical stages	41
Feed tray	21
Feed flow rata F	1.0 (kmol/min)
Top product rata D	0.5 (kmol/min)
Bottom product rate B	0.5 (kmol/min)
Reflux rate L	2.7063 (kmol/min)
Steam rate V	3.2063 (kmol/min)
Feed composition	0.5
Top composition x_D	0.99
Bottom composition x_B	0.01

2.2 Dynamic simulation conditions and process data collection

The data sets needed to develop the composition estimators of the distillation column were generated by running the first-principles dynamic model under these operating conditions.

To generate data for building the composition estimator models, random perturbations within ±10% of the steady-state value were added to the feed composition during simulations. In addition to these random disturbances, the total feed flow rate changes stepwise by ±1% every 2 h, while the fluctuation of the total flow rate is restricted within ±2% of its steady-state value. Measurement noises of the distribution $N(0^oC, 0.1^oC)$ were added to the tray temperatures measurements. Simulated data for validating composition estimators are obtained under the almost same conditions as described above. The differences are the seeds of the random signals. The sampling period of the compositions as well as other process variables is set at 1 min. The total simulation time is set at 6 h.

By this way, we collected 360 sampling data as the training and testing data sets. The first 300 data were used as training data while the last 60 data were used as testing data. Fig.2 shows the 360 data of top product compositions.

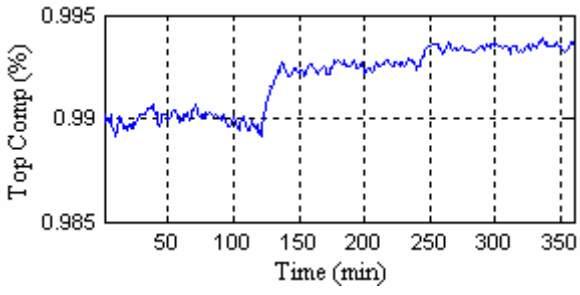

Fig 2. Top product compositions

3. SENSITIVITY MATRIX ANALYSIS

The choice of the secondary variables is one of the key technique of composition estimators. Proper secondary variables selection can make the estimator be built based on the right relation of the input and output sample data. In this paper, we choose the most sensitive secondary variable from the sensitivity matrix directly and application it for composition estimators of the continuous distillation column.

The sensitivity matrix analysis is described as follows. The sensitivity matrix is defined as the partial derivative of each secondary variable with respect to each primary variable to be estimated. The sensitivity matrix calculated for all the available process variables are collected in a gain matrix S :

$$S=\begin{bmatrix} \frac{\partial T_1}{\partial x_1} & \cdots & \frac{\partial T_1}{\partial x_i} & \cdots & \frac{\partial T_1}{\partial x_m} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial T_j}{\partial x_1} & \cdots & \frac{\partial T_j}{\partial x_i} & \cdots & \frac{\partial T_j}{\partial x_m} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial T_n}{\partial x_1} & \cdots & \frac{\partial T_n}{\partial x_i} & \cdots & \frac{\partial T_n}{\partial x_m} \end{bmatrix}^T \tag{2}$$

Where T_j is the j th secondary variable, x_i represents the i th primary variables, n is the number of available secondary variable, and m is the number of primary variable to be estimated.

The $m\times n$ sensitivity matrix S can be determined from simulations based on a first-principles process model. In this paper, the sensitivity matrix is calculated by the following approximation:

$$\hat{S}=\begin{bmatrix} \frac{\Delta T_1}{\Delta x_1} & \cdots & \frac{\Delta T_1}{\Delta x_i} & \cdots & \frac{\Delta T_1}{\Delta x_m} \\ \vdots & & \vdots & & \vdots \\ \frac{\Delta T_j}{\Delta x_1} & \cdots & \frac{\Delta T_j}{\Delta x_i} & \cdots & \frac{\Delta T_j}{\Delta x_m} \\ \vdots & & \vdots & & \vdots \\ \frac{\Delta T_n}{\Delta x_1} & \cdots & \frac{\Delta T_n}{\Delta x_i} & \cdots & \frac{\Delta T_n}{\Delta x_m} \end{bmatrix}^T \tag{3}$$

Where $\Delta x_i = x_i(t+\Delta t)-x_i(t)$ indicates the variation of the i th primary variable during the selection time interval Δt, and $\Delta T_j = T_j(t+\Delta t)-T_j(t)$ represents the variation of the j th secondary variable in the same period Δt.

In order to get the most sensitive secondary variable directly in this matrix $\hat{S}$, the absolute value of every element is used:

$$\tilde{S}=\left|\hat{S}\right| \tag{4}$$

The sensitivity matrix analysis method to choose the optimal secondary variables directly as the composition estimator inputs is described as follows. Because the largest value $\tilde{s}_{i,j}$ in the sensitivity matrix $\tilde{S}$ is a measure of the sensitivity of the j th secondary variable, the secondary variable having the largest value of $\tilde{s}_{i,j}$ could be considered as the most suitable estimator input. Similarly, the location having the secondary largest value of $\tilde{s}_{i,j}$ is the second most appropriate estimator input, and so on.

4. KERNEL RIDGE REGRESSION ALGORITHM

The kernel ridge regression algorithm is described as follows.

Suppose we have a training set $S=\{x_i,y_i\}_{i=1}^{l}$, where $x_i\in X\subseteq\mathbb{R}^n$, $y_i\in Y\subseteq\mathbb{R}$, and $i=1,\ldots,l$.

Assume ϕ is a nonlinear mapping ($\phi:\mathbb{R}^n\to F$), which transform the vectors in the input space into vectors in some high dimensional feature space, where $F=\{\phi(x)\,|\,x\in X\}$.

$$x=(x_1,\cdots x_l)\to\phi(x)=(\phi_1(x),\cdots\phi_l(x)) \tag{5}$$

In the high dimensional feature space F, the goal is to estimate a model of this form

$$y=\sum_{i=1}^{l}w_i\phi_i(x)+b=\langle w\cdot\phi(x)\rangle+b \tag{6}$$

The optimization problem is defined as this:

$$L(w)=\lambda\langle w\cdot w\rangle+\sum_{i=1}^{l}(y_i-\langle w\cdot\phi(x_i)\rangle-b)^2 \tag{7}$$

Where $\lambda\langle w\cdot w\rangle$ is the regularization term and λ is the regularization parameter.

Assume that $b=0$, so we can express (7) as follows:

$$L(w)=\lambda w'w+(y-Xw)'(y-Xw) \tag{8}$$

Where $X=\begin{pmatrix}\phi(x_1)'\\ \phi(x_2)'\\ \vdots\\ \phi(x_l)'\end{pmatrix}$, $y=\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_l\end{pmatrix}$.

Differentiating (8) in w, we obtain

$$\frac{\partial L}{\partial w}=0\Rightarrow w=(X'X+\lambda I_l)^{-1}X'y \tag{9}$$

With the application of Mercer's theorem on the kernel matrix (Vapnik, 1999) $K_{ij}=K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$, it is not required to compute explicitly the nonlinear mapping ϕ as this is done implicitly through the use of positive definite kernel functions K.

By elimination of w, the following nonlinear kernel ridge regression model is obtained as

$$y = f(x) = \langle w \cdot \phi(x_i) \rangle = y'(K + \lambda I_l)^{-1} k \quad (10)$$

There are some typical kernel functions (Vapnik, 1999):

1) Polynomial kernels: $K(x, x_k) = [(xx_k + 1]^d$;

2) Radial basic function kernels:

$$K(x, x_k) = \exp\left(- \frac{\| x - x_k \|}{2 p^2} \right);$$

3) Network kernels: $K(x, x_k) = S(v(x, x_k) + c)$.

5. COMPOSITION ESTIMATOR

5.1 Optimal selection of estimator inputs

In order to gain the sensitivity matrix $\tilde{S}$, firstly, we should get the (3). In this paper, the sample time interval Δt is defined as $\Delta t = 1$ s. In section 2.2, the sampling period of the compositions as well as other process variables is set at 1 min. That is to say, we also collect the sample data with the sampling period at 61 s. Computing every element $\Delta T_j / \Delta x_i$ of (3) and the absolute value of every element, we get the sensitivity matrix $\tilde{S}$.

In this distillation column, the temperatures on 41 trays are used as the secondary variables and the top product composition is used as the primary variable. We can directly analysis the matrix $\tilde{S}$ to get the optimal secondary variables.

Fig. 3 shows the locations of the most sensitive temperatures identified from the matrix $\tilde{S}$.

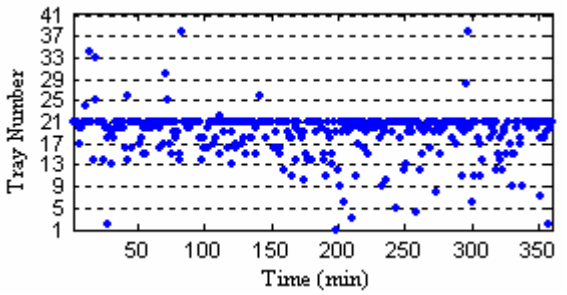

Fig 3. The most sensitive input

Because the largest value $\tilde{s}_{i,j}$ in the sensitivity matrix $\tilde{S}$ is a measure of the sensitivity of the j th secondary variable, the secondary variable having the largest value of $\tilde{s}_{i,j}$ could be considered as the most suitable estimator input. We can select out the location of the largest value from the matrix $\tilde{S}$ and the Fig. 3 shows the most sensitive location with the sample time. We can see that the most sensitive temperature changes with the sample time. But in Fig.3, we can find that the frequency of temperature on the 21 tray appears more than the others. So the temperature on the 21 tray is considered as the most sensitive variable to the top composition.

Moreover, Fig. 4, Fig. 5, Fig. 6 shows the 2nd, 3rd and 4th most sensitive tray to the top composition. The same as we choose the most sensitive tray from Fig. 3, the temperatures on 20, 22, 19 trays are chosen as the 2nd, 3rd and 4th most sensitive tray to the top composition.

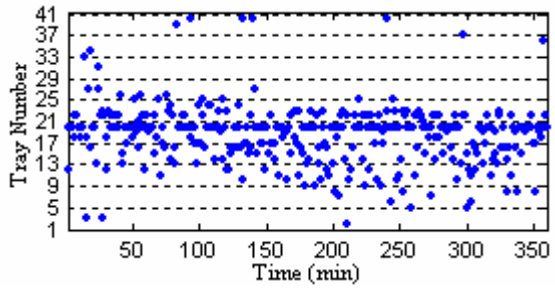

Fig 4. The 2nd most sensitive input

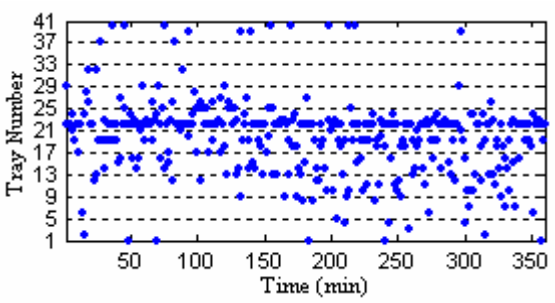

Fig 5. The 3rd most sensitive input

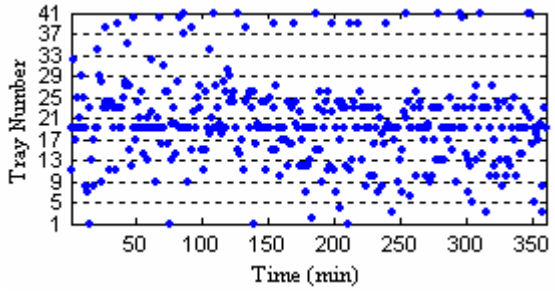

Fig 6. The 4th most sensitive input

Quintero, Luyben and Georgakis (1991) suggested that $N_c + 2$ temperature measurements should be considered, where N_c is the number of chemical components in the feed flow. Because the components $N_c = 2$ in this paper, followed them, the temperatures on 21, 20, 22, 19 trays are chosen as the sensitive variables to the composition estimator inputs.

5.2 Development of composition estimator using kernel ridge regression

In this paper, the radial basic function kernel is chosen as the kernel function:

$$K(x, x_k) = \exp\left(-\frac{\|x - x_k\|}{2p^2}\right)$$

The optimal selection of kernel parameter p and regularization parameter λ are very important to the KRR estimator. In this paper, the cross-validation technique is used.

The composition estimator is evaluated on the basis of root mean squared error (RMSE) of prediction, which is calculated by applying the model to the validation data,

$$RMSE = \left(\frac{1}{N}\sum_{n=1}^{N}(x(n)-\hat{x}(n))^2\right)^{\frac{1}{2}} \quad (11)$$

where x is a measurement of product composition, $\hat{x}$ is its prediction, and N is the number of measurements.

As a result, the parameters of KRR estimator are defined as $p = 2.7$, $\lambda = 1\times10^{-7}$. The prediction results of KRR estimator for training data and the prediction results for testing data are shown in Fig. 7. As a comparison, the prediction results of RBF neural networks estimator for training data and the prediction results for testing data are shown in Fig. 8. In the Fig. 7 and Fig. 8, the solid lines represent the true simulated distillation compositions while the dashed lines represent the KRR model predictions and the RBF model predictions. The statistical results of the two models in the testing data set are summarized in Table 2.

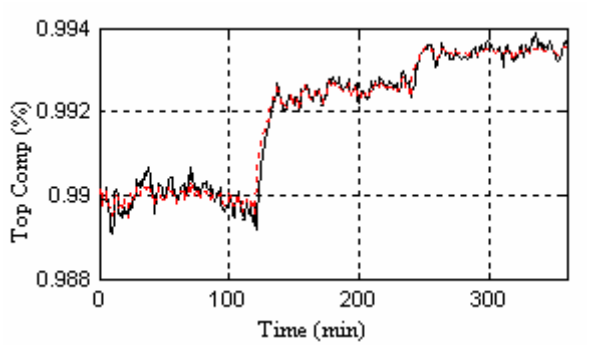

Fig 7. Prediction from the KRR estimator

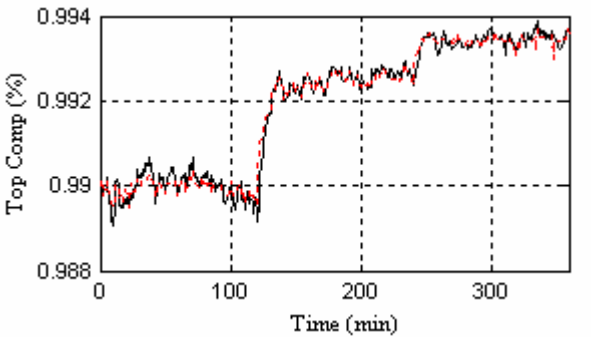

Fig 8. Prediction from the RBF estimator

Table 2. Comparison of composition estimators

Model	RMSE (10^{-4})
RBF	1.4680
KRR	1.0371

From Fig. 7, it is found that the KRR estimator has good performance in estimation of the top composition. Estimated outputs of estimator based on KRR to the top composition match real values of the top composition and follow the varying trend of the top composition very well. From Fig. 8, we also find that RBF estimator has the good performance in the first 300 training data, but in the last 60 testing data its prediction performance is bad. From Table 2, it is shown that KRR estimator has good performance of generalization. Compared with the RBF estimator, statistical results in Table 2 show that the KRR estimator predicts more accurate in the last 60 testing data.

5.3 Influence of measurement noise

In subsection 5.1 and 5.2, the noise added to the temperature measurement is $N(0^oC, 0.1^oC)$, the variance $\sigma^2 = 0.1^oC$. In order to consider the effect of measurement noise to the estimator, in this subsection, the variance $\sigma^2 = 0.5^oC$ of the Gaussian noise is added to the temperature variables. The other dynamic conditions are the same as subsection 2.2. The parameters of KRR estimator are chosen as $p = 2.7$, $\lambda = 1\times10^{-7}$. Fig. 9 shows the prediction results from the KRR and RBF estimator. The statistical results of the two models in the testing data set under this high-level noise are summarized in Table 3.

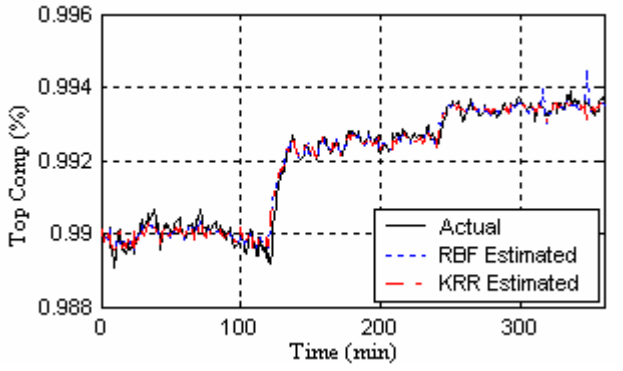

Fig 9. Prediction from the KRR and RBF estimator

Table 3. Comparison of composition estimators

Model	RMSE (10^{-4})
RBF	1.7235
KRR	1.2795

From Fig. 9, we also can see that the KRR estimator of the top composition match actual values of the top composition and follow the varying trend of the top composition very well. From Table 3, it is shown that the performance is a little deteriorated by measurement noise. However, it seems to be acceptable, the KRR estimator has good performance of generalization. Furthermore, the influence of measurement noise can be suppressed by filtering or scaling data appropriately. The simulation results demonstrate that the estimator based on the sensitivity matrix analysis and KRR is effective and robust.

6. CONCLUSIONS

This paper does research on estimating of distillation composition on-line based on a novel method of sensitive matrix analysis and kernel ridge regression. Through the optimal choice of the second variables and building the KRR composition model, the simulation result shows that this method is efficient.

With the development of composition estimators, our work makes it possible to implement advanced control of quality variables of the distillation process on-line. In summary, successful installation of the composition estimators in an existing refinery can ensure better product quality control with higher productivity.

7. ACKNOWLEDGEMENT

This work was supported by National Basic Research Program of China under grant No. 2007CB714006.

REFERENCES

Kano, M., Miyazaki, K., Hasebe, S., and Hashimoto, I. (2000). Inferential control system of distillation compositions using dynamic partial least squares regression. *Journal of Process Control*, **10**, 157-166.

Kano, M., Miyazaki, K., Hasebe, S., and Hashimoto, I. (2003). Inferential control of distillation compositions: selection of model and control configuration. *Control Engineering Practice*, **11**, 927-933.

Quintero, M, E., Luyben, W, L., and Georgakis, C. (1991). Application of an Extender Luenberger observer to the control of multicomponent batch distillation. *Industrial and Engineering Chemistry Research*, **30**, 1870-1880.

Joseph, B., and Brosilow, C, B. (1978). Inferential control of processes: Part I. Steady state analysis and design. *AIChE Journal*, 1978, **24**, 485-492.

Yan, W, W., Shao, H, H., and Wang, X, F. (2004) Soft sensing modeling based on support vector machine and Bayesian model selection. *Computers and Chemical Engineering*, **28**, 1489-1498.

Bhartiya, S., and Whiteley, J, R. (2001). Development of inferential measurements using neural networks. *ISA Transactions*, **40**, 307-323.

Mejdell, T., and Skogestad, S. (1991a). Estimation of distillation compositions from multiple temperature measurements using partial-least-squares regression. *Industrial and Engineering Chemistry Research*, **30**, 2543-2555.

Mejdell, T., and Skogestad, S. (1991b). Composition estimator in a pilot-plant distillation column using multiple temperatures. *Industrial and Engineering Chemistry Research*, **30**, 2555-2564.

Jie Zhang. (2001). Inferential feedback control of distillation composition based on PCR and PLS models. *Proceedings of the American Control Conference*, 1196-1201.

Zamprogna, E., Barolo, M., and Seborg, D, E. (2004). Estimating product composition profiles in batch distillation via partial least squares regression. *Control Engineering Practice*, **12**, 917-929

Zamprogna, E., Barolo, M., and Seborg, D, E. (2005). Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis, *Journal of Process Control*, **15**, 39-52.

Vapnik, V. (1999). *The nature of statistical learning theory (2nd ed)*. Springer, Berlin.

Hoerl, AE., and Kennard, RW. (1970a). Ridge Regression: Biased Estimation for Non-orthogonal Problems. *Technometrics,* **12**, 55-67.

Hoerl, AE., and Kennard, RW. (1970b). Ridge Regression: Application for Non-orthogonal Problems. *Technometrics* , **12**, 69-72.

Saunders, C., Gammerman, A., and Vovk, V. (1998). Ridge regression learning algorithm in dual variables. *Proceedings of the 15th International Conference on Machine Learning*, 515-521.