8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.3, June 6-8, 2007, Cancún, Mexico

INFERRING DISTILLATION PRODUCT COMPOSITION: A HYBRID SOFT SENSOR APPROACH

I.Y. Smets [*] **S. Boon** [*] **T. Boelen** [***] **J. Espinosa** [***,****]
J.F. Van Impe [*,**]

[*] BioTeC, *Dept. of Chemical Engineering, K.U.Leuven, W. de Croylaan 46, B-3001 Leuven, Belgium*
[**] *Corresponding author: jan.vanimpe@cit.kuleuven.be*
[***] IPCOS, *Technologielaan 11-0101, B-3001 Leuven, Belgium*
[****] *Faculty of Mines, Universidad Nacional de Colombia, Cra.80 No.65-223, Medellín, Colombia*

Abstract: Adequate process monitoring and optimal control of distillation columns relies heavily on accurate and preferably on-line estimates of the product composition. Hence, inferring the product composition from easily accessible and abundantly available process measurements has become a key element for successful operation. This paper compares a hybrid soft sensor approach, based on the General Distillation Shortcut method introduced by Friedman in 1995, with a pure black box approach. On the basis of two industrial multicomponent distillation case studies, it can be concluded that the hybrid GDS approach outperforms the black box one if (*i*) a temperature measurement is available that is *sensitive* for the to be predicted concentration and (*ii*) if that concentration is present in a *substantial* amount with respect to the other components. The black box soft sensors do not suffer from that last drawback but, once again, their lack of extrapolative power is clearly illustrated.

Keywords: distillation columns, soft sensing, monitoring, hybrid modeling, black box modeling, industrial case study

1. INTRODUCTION

For adequate process monitoring and control, an accurate estimation of the product compositions during distillation is a prerequisite. Although product composition can be measured online, most analyzers, like gas chromatographs and NIR (Near-Infrared) analyzers are expensive and difficult to maintain. Furthermore, they entail significant measurement delays precluding *in time* control actions.
Hence, inferring the product composition from easily accessible and abundantly available process measurements has become a key element for successful operation. The development of such *inferential* or *soft sensor* controllers is far from new but remains highly relevant as witnessed by recent publications in this domain. The type of model on which the soft sensors rely, varies from first principles, mechanistic models to black box models.
As correctly formulated by Kano *et al.* (2000) a first principles model is preferred as far as it is available and provides sufficient accuracy with reasonable computational load. Predominantly, strategies based on Extended Kalman Filters are proposed (e.g., (Baratti *et al.*, 1995; Baratti *et al.*, 1997; Lee and Morari, 1992; Oisiovici and Cruz, 2001)), but, more recently, also a combination of the wave propagation equation with static mass and energy balances has been reported (Roffel *et al.*, 2003).
If, however, no fundamental model appropriate for real-time use exists, an empirical model must be derived from process data. With the huge amount of data that is nowadays stored in computers, this black box modeling represents a feasible challenge. At present, techniques based on multivariate statistics such as *Partial Least Squares* (PLS)

are very popular, mainly since they can handle the collinearity in the data very well. Numerous applications have been described for continuous, distillations (see, e.g., (Kamohara *et al.*, 2004; Kano *et al.*, 2000; Kano *et al.*, 2003; Kresta *et al.*, 1994; Mejdell and Skogestad, 1991*a*; Mejdell and Skogestad, 1991*b*; Pannocchia and Brambilla, 2003; Park and Han, 2000; Shin *et al.*, 1999)) as well as batch distillations (e.g., (Zamprogna *et al.*, 2004)).
Apart from these statistics based black box techniques, neural networks, enabling *nonlinear* modeling of the various relationships, are definitely still in the running (e.g., (Bahar *et al.*, 2004; Baratti *et al.*, 1997; Yeh *et al.*, 2003)).
Unfortunately, no need for any fundamental process knowledge to build the models, implies the well known lack of extrapolative power. The latter becomes prominent when one has to build a soft sensor based on time series of regular, day to day, *available* process data.
The motivation of the here presented research was, therefore, to check whether the incorporation of a limited amount of fundamental knowledge can improve the extrapolability and, hence, the quality of the composition estimators. This hybrid approach is compared with a pure black box approach. Given the cost of dedicated experiments in an industrial setting, the training of both types of models has to be performed on available data sets retrieved from regular operating conditions. Hereto, two industrial multicomponent distillation case studies will be considered.
The structure of the paper is as follows. In Sections 2 and 3 the hybrid and the black box model based soft sensors are introduced, respectively. Then, in Section 4, the two industrial case studies are discussed. Section 5 presents the obtained results and Section 6 summarizes the main conclusions.

2. HYBRID SOFT SENOR BASED ON SHORTCUT FIRST PRINCIPLES MODELS

While rigorous models based on the MESH equations, are too involved to be implemented online, a *shortcut* first principles model will be the starting point for the here proposed hybrid soft sensor.

The General Distillation Shortcut (GDS) method was developed by Friedman (1995) and involves a short cut simulation of a section of the column, typically a bottom half of the stripping section or a top half of the rectifying section (Friedman *et al.*, 2002). The GDS method relies on Colburn's formulae for distillation column section performance (Colburn, 1941), describing the ratio between tray composition and bottom (or top) composition as a function of the components' volatility (K value), internal reflux, and number of trays in the section. More specifically, the following four equations are involved.

(1) **Bubble point equation**:

$$\sum_i K_i x_i = 1 \tag{1}$$

with K_i the equilibrium constant between the fluid fraction x_i and vapor fraction y_i of component i.

(2) **Dew point equation**:

$$\sum_i \frac{x_i}{K_i} = 1 \tag{2}$$

Normally, this equation should be written in terms of the vapor fraction y_i instead of the fluid fraction x_i but here it is assumed that the vapor fraction y_i at a certain position in the column equals the fluid fraction x_i somewhere else in the column.

(3) **Colburn relation**:

$$\sum_i x_i R_i = 1 \tag{3}$$

In the original contributions of Friedman, the R-factor in the Colburn relation is only provided for the bottom section (for brevity, subscript i referring to component i is omitted), i.e.,

$$R_{bottom} = \frac{U^{N+1} - 1}{U - 1}(K - 1) + 1$$

with U [-] being equal to $\frac{K \cdot V'}{L'}$ in which V' [mole/hr] represents the internal vapor stream in the bottom section and L' [mole/hr] the internal liquid stream in the bottom section. N is the number of plates up to the *sensitive* plate (starting from the bottom).

A similar expression has here been derived for the R-factor of the top section:

$$R_{top} = \frac{(\frac{1}{U})^{N+1} - 1}{\frac{1}{U} - 1}(\frac{1}{K} - 1) + 1$$

with U [-] being equal to $\frac{K \cdot V}{L}$ in which V [mole/hr] represents the internal vapor stream in the top section and L [mole/hr] the internal liquid stream in the top section. N is the number of plates up to the *sensitive* plate (starting from the top).

(4) **Sum equation**:

$$\sum_i x_i = 1 \tag{4}$$

A number of variables remain to be specified such as the equilibrium constant K and the internal vapor and liquid streams.

Equilibrium constant K. Since fugacity and activity coefficients are hard to obtain for a mixture, also the equilibrium constant K_i has to be approximated. Hereto, the ratio of the vapor pressure P^{vap} over the column pressure P turned out to be more reliable than the approximation based on the internal vapor and liquid streams (for the heavy and light key component). Hence, the equilibrium constant K_i is implemented as

$$K_i = \frac{P^{vap}}{P}$$

in which the vapor pressure P^{vap} is approximated by the relation of Antoine

$$\ln(P^{vap}) = A - \frac{B}{T + C}$$

In this work, the A, B and C coefficients are taken from (Prausnitz, 1977).

Internal vapor and liquid streams. For the calculation of the internal vapor and liquid streams, the assumption of a *constant molar overflow* is adopted which is fairly valid for components from the same homologous series. The internal flows, i.e., $V^{'}$ and $L^{'}$ in the top section and V and L in the bottom section, can then be calculated as described below

$$\begin{aligned}
V^{'} &= \frac{Q_R}{\Delta H_v} \\
V &= V^{'} - F\frac{c_F}{\Delta H_v}(T_F^b - T_F) = V^{'} - (q-1)F \\
L &= R(1 + \frac{c_R}{\Delta H_v}(T_R^b - T_R)) \\
L^{'} &= L + F(1 + \frac{c_F}{\Delta H_v}(T_F^b - T_F)) = L + qF
\end{aligned}$$

with

$$q = 1 + \frac{c_F}{\Delta H_v (T_F^b - T_F)}$$

and
Q_R [J/hr] being the reboiler duty, ΔH_v [J/mole] the vaporization heat, F [mole/hr] the feed flow, R [mole/hr] the reflux flow, c_F [J/mole/K] the heat capacity at constant pressure of the feed, c_R [J/mole/K] the heat capacity at constant pressure of the reflux stream, T_F [K] the temperature of the feed, T_R [K] the temperature of the reflux stream, T_F^b [K] the boiling temperature of the feed and T_R^b [K] the boiling temperature of the reflux stream.

Reboiler duty. Given the general design equation for the reboiler duty

$$Q_R = (H^{vap} + c_{p,steam} \cdot (T_{in,steam} - T_{out,steam})) \cdot F$$

the observer will rely on the equation

$$Q_R = k \cdot F$$

with

$$k = H^{vap} + c_{p,steam} \cdot (T_{in,steam} - T_{out,steam})$$

Since the variation in reboiler duty is dominated by the variations in feed flow, it has been tested whether k can be approximated by a constant value, inferred from the data.

A considerable advantage of the GDS method is that the composition of the feed does not have to be known. Disadvantage is that, by solving four equations, the method seems only appropriate for a mixture of maximum four components. While Friedman himself has tested this hybrid approach (to which he refers to as a first principles approach) already on several industrial case studies, with results ranging from acceptable to excellent, the performance was never compared with the nowadays more popular black box approaches.

3. BLACK BOX SOFT SENSORS

The black box model based soft sensors are built with the `PRESTO` tool (IPCOS (Belgium)). Since collinearity of the data is reflected as a lack of excitation which limits the accuracy and increases the sensitivity of the model parameters, `PRESTO`, as a tool, applies a Partial Least Squares (PLS) transformation on the plant data in order to obtain the main directions of variability of the plant which are correlated with the estimated variable. It is up to the user to decide the number of directions (number of components) taken into account to build the model.
This transformation is used directly to construct linear static models, since they are linear regressions over the input parameters using the PLS method. When the process data used for soft sensor design exhibits dominant dynamic effects, mainly caused by mixing effects on the plates, reflux drum or bottom of the column, state space models are exploited. The user is able to define the order of these state space models.
In addition, `PRESTO` can estimate nonlinear static models and nonlinear dynamic models. For his purpose the user can define a special type of fuzzy model called `GNOMO` (Generalized NOnlinear Model) in `PRESTO`. This model has very interesting features including monotonicity and limited nonlinear mapping which are very important to guarantee the safe and robust operation of the soft sensor. The user can define the complexity of the model and the degree of nonlinearity.

4. INDUSTRIAL CASE STUDIES

Two industrial case studies are considered. For the first case study, referred to as the LPG plant, a total of three soft sensors has to be developed, i.e., a sensor for the *residual* concentration of ethane (C2) at the bottom of the first column, a sensor for *residual* propane (C3) in the bottom of the second column and a sensor for *residual* butane (C4) in the top of the second column. Historical data (comprising, e.g., temperature and pressure information) of five months are available with a sampling rate of one minute. GC analyzer data for the composition of the top and bottom streams is available every 30 minutes (with spline interpolation for data alignment). Although a lot of data is available, the data of at most one month will be used as training data (and the remaining months as validation data) since it is advantageous if the soft sensors can be built on a restricted historical data set.

The second case study deals with a styrene (SM) production plant. Again three soft sensors have to be developed for (the first) two distillation columns of the treatment train: a soft sensor is needed for (i) the top ethylbenzene concentration (EB) in the first column (pure ethylbenzene is recycled to the reaction part of the plant), the bottom toluene concentration (TOL) for the first column and (iii) the bottom ethylbenzene concentration in the second column. The available historical data is spread over one month of one year and six weeks in the next year. The former will be used as training data while the latter will serve as validation data. Also here GC analyzer data is available (sampling periods of 15 minutes).

5. RESULTS

5.1 *Optimization criterion*

To identify the missing parameter values in the hybrid soft sensors the root mean square criterion is adopted. The performance of the resulting models will, however, be graphically illustrated.

5.2 *LPG case study*

Due to space limitations, only the first soft sensor will be discussed, the performance of the remaining two soft sensors being comparable.

5.2.1. *GDS soft sensor: C2 at the bottom of the first column.* Since a temperature measurement before as well as after the reboiler is present, all four GDS equations can be exploited. Only one extra temperature in the bottom section is available (plate 3) turning this temperature automatically into the *sensitive* one. As can be seen from Figure 1, with training data of one week, quite accurate results (grey line) when compared with the analyzer data (black line) can be obtained for the component that has to meet a certain specification there.

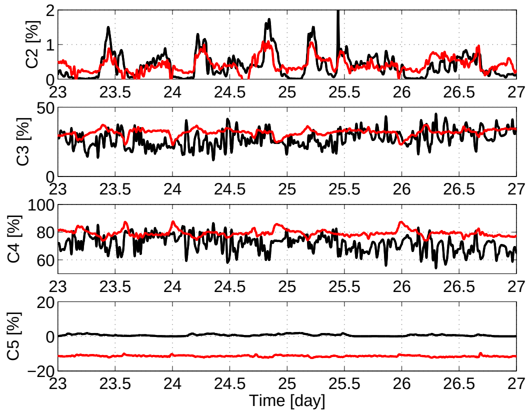

Fig. 1. LPG case study: training results (one week) of the GDS C2 sensor (grey line). Analyzer measurements: black line.

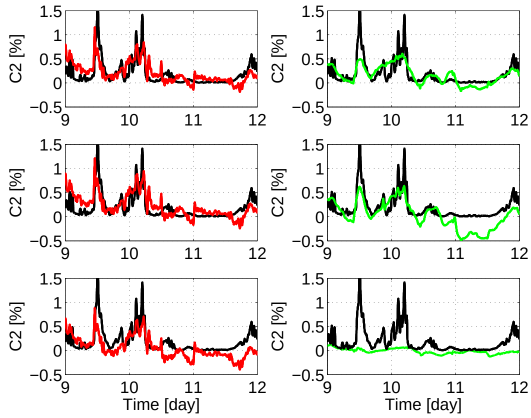

Fig. 2. LPG case study: validation results of the GDS (left) and `PRESTO` (right) C2 sensor. Model trained with 1 month (grey line, top), 1 week (grey line, center) and 1 day data (grey line, bottom). Analyzer measurements: black line.

The other components, especially the butane (C4) and pentane (C5) concentrations, are less well predicted since the selected *sensitive* temperature is probably not sensitive for these components at that location in the column.

Validation on the data of the remaining months indicates a robust performance of the soft sensor for the ethane (C2) concentration prediction. Moreover, even validation with a model that is trained on one day data, performs satisfactorily as can be seen from Figure 2 (left).

5.2.2. `PRESTO` *soft sensors.* As mentioned before, with the `PRESTO` tool, several types of models can be built. In this study the focus was on static linear, dynamic linear and static nonlinear models. With respect to the *optimal* type of soft sensor, no clear conclusion can be drawn. A representative result is depicted in Figure 2 (right) for the C2 sensor at the bottom of the first column. When comparing the performance of the black box soft sensors with the hybrid GDS sensors,

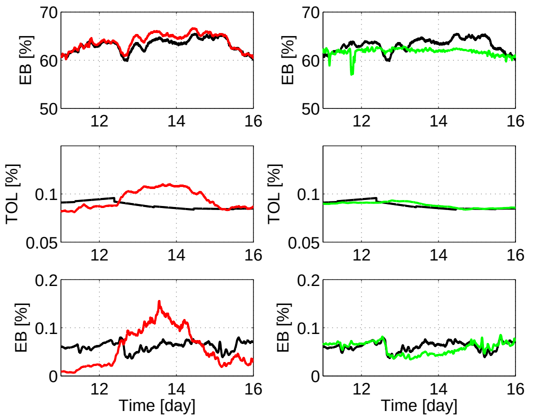

Fig. 3. SM case study: training results of the GDS (grey line, left) and `PRESTO` (grey line, right) top EB concentration (top), bottom TOL concentration (center) and bottom EB concentration (bottom) sensors. Models trained with 1 month data. Analyzer measurements: black line.

it can concluded that the GDS sensors are more robust since their validation performance does not deteriorate significantly when the training data set reduces from one month, over one week to one day which cannot be said of the `PRESTO` soft sensors. This statement holds also for the other soft sensors of this case study.

5.3 *Styrene case study*

5.3.1. *GDS soft sensor: EB at the top of the first column.* Since the feed contains 22 components of which only four can be estimated, only the three dominant species are considered, i.e., toluene, ethylbenzene and styrene while the other components are grouped into a fictitious rest component. Hence, only three GDS equations will be considered: the dew point equation in the top, the Colburn equation and the sum equation. The latter does not sum up to one but to (1-x_{rest}), a value that is identified from the data. As illustrated in Figure 3 (top left) the performance of the soft sensor with one month training is very satisfying for ethylbenzene.

In validation, the performance of the soft sensor is corroborated (Figure 4 (left)).

5.3.2. *GDS soft sensor: TOL at the bottom of the first column.* While the relative concentration of ethylbenzene at the top of this column approximates 65%, the to be predicted toluene concentration at the bottom is much smaller (0.089%). Not taking into account the concentration of the 19 other components has, therefore, a highly negative impact on the performance (Figure 3 (center left)).

5.3.3. *GDS soft sensor: EB at the bottom of the second column.* Again, only the three dominant

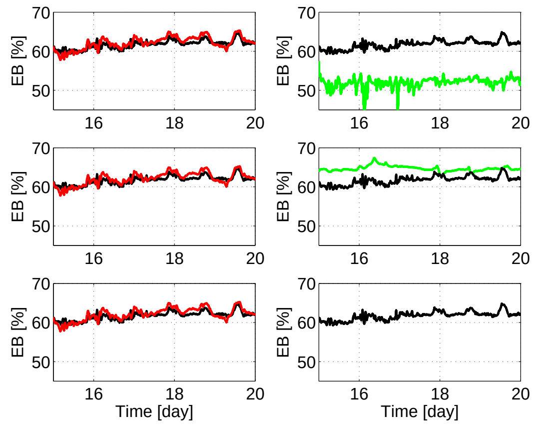

Fig. 4. SM case study: validation results of the GDS (grey line, left) and `PRESTO` (grey line, right) top EB concentration. Models trained with 1 month (top), 1 week (center) and 1 day (bottom) data. Analyzer measurements: black line.

components are taken into account, i.e., ethylbenzene, α-methyl styrene and styrene. Unfortunately, also here, the to be predicted ethylbenzene concentration at the bottom is very low, precluding a proper estimation (Figure 3 (bottom left)).

5.3.4. `PRESTO` *soft sensors.* For the EB sensor at the top of the first column, results are compared between one month, one week and one day training. For the remaining soft sensors, one month of training is imposed. As evidenced by Figure 3 (right), the black box soft sensors (the best of all investigated types is shown) perform much better for the cases where the GDS soft sensors failed, i.e., at very low concentrations of the component under consideration with respect to the other concentrations. If the latter's concentration is, however, high enough with respect to the other components, the GDS soft sensor performs better. Turning our attention towards validation, it is clear that, if a well performing GDS soft sensor is obtained during training, this soft sensor is quite robust since the validation (even if only trained with one day data) remains of high quality, while in this case study the validation of the black box soft sensors fails completely (Figure 4 (right)). The model that was trained with one day data performs so badly that the prediction does not even fit the figure window (bottom right plot). Reason is the well known lack of extrapolative power. Indeed, the validation data set is quite different from the training data set since the top pressure was changed (from 160 to 137 mbar) inducing different flows, refluxes and temperature profiles (e.g., a top temperature of 66°C instead of orginally 71°C).

6. DISCUSSION AND CONCLUSIONS

In this paper, a hybrid soft sensor, based on the General Distillation Shortcut method introduced by Friedman, is proposed and compared with pure black box soft sensors to infer the composition of the top and bottom flows in multicomponent distillation columns. The GDS soft sensor exploits the bubble point, dew point, Colburn and sum equations while different types of black box models (static linear, dynamic linear as well as static nonlinear) are tested.

As evidenced in this work, the incorporation of a limited amount of *mechanistic* knowledge (or process specific knowledge) improves the robustness of the sensor.

Since the GDS soft sensor is based on four equations, at most four concentrations can be estimated. If the mixture contains more than four components, the dominant components should be selected while the other components must be grouped in one *rest* component. Following this procedure (for the second case study), it can be concluded that the GDS soft sensor does not perform well if the to be predicted concentration is present in very low concentrations. It is, however, unclear at this moment whether this is due to the large amount of rest components or to the fact that these rest components are not known. The black box soft sensors do not suffer from this drawback and perform very well in training. If, however, the working conditions of the plant change, the validation performance of the black box soft sensors is extremely low.

Hence, in general, if the (relative) concentration of the to be predicted component is not too small and if there is a temperature available that is sensitive to it, the GDS soft sensor exhibits a high quality performance. In addition, the required training data set is very limited (one week or even one day) while no dedicated experiments are needed.

ACKNOWLEDGMENTS

Work supported by Projects OT/03/30 and EF/05/006 (Center of Excellence - Optimization in Engineering (OPTEC)) of the K.U.Leuven Research Council, and the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science Policy Office. The scientific responsibility is assumed by its authors.

7. REFERENCES

Bahar, A., C. Özgen, K. Leblebicioglu and U. Halici (2004). Artificial neural network estimator design for the inferential model predictive control of an industrial distillation column. *Industrial Engineering Chemistry Research* **43**, 6201–6111.

Baratti, R., A. Bertucco, A. Da Rold and M. Morbidelli (1995). Development of a composition estimator for binary distillation columns. Application to a pilot plant. *Chemical Engineering Science* **50**(10), 1541–1550.

Baratti, R., S. Corti and A. Servida (1997). A feedforward control strategy for distillation columns. *Artificial Intelligence in Engineering* **11**, 405–412.

Colburn, A.P. (1941). *Industrial Engineering Chemistry Research* **33**, 495.

Friedman, Y.Z. (1995). Simulation based inferential control. *AIChE spring meeting* http://petrocontrol.com/GDS.html.

Friedman, Y.Z., E.A. Neto and C.R. Porfirio (2002). First-principles distillation inference models for product quality prediction. *Hydrocarbon processing journal,* **February**, http://petrocontrol.com/GDS.html.

Kamohara, H., A. Takinami, M. Takeda, M. Kano, S. Hasebe and I. Hashimoto (2004). Product quality estimation and operation condition monitoring for industrial ethylene fractionator. *Journal of Chemical Engineering of Japan* **37**(3), 422–428.

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

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

Kresta, J.V., T.E. Marlin and J.F. MacGregor (1994). Development of inferential process models using PLS. *Computers and Chemical Engineering* **18**(7), 597–611.

Lee, J.H. and M. Morari (1992). Robust inferential control of multi-rate sampled-data systems. *Chemical Engineering Science* **47**(4), 865–885.

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

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

Oisiovici, R.M. and S.L. Cruz (2001). Inferential control of high-purity multicomponent batch distillation columns using an Extended Kalman Filter. *Industrial Engineering Chemistry Research* **40**, 2628–2639.

Pannocchia, G. and A. Brambilla (2003). Consistency of property estimators in multicomponent distillation control. *Industrial Engineering Chemistry Research* **42**, 4452–4460.

Park, S. and C. Han (2000). A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns. *Computers and Chemical Engineering* **24**, 871–877.

Prausnitz (1977). *The properties of gases and liquids.* McGraw-Hill Book Company (New York).

Roffel, B., B.H.L. Betlem and R.M. de Blouw (2003). A comparison of the performance of profile position and composition estimators for quality control in binary distillation. *Control Engineering Practice* **27**, 917–929.

Shin, J., M. Lee and S. Park (1999). Design of a composition estimator for inferential control of distillation columns. *Chemical Engineering Communications* **178**, 221–248.

Yeh, T.-M., M.-C. Huang and C.-T. Huang (2003). Esimate of process compositions and plantwide control from multiple secondary measurements using artificial neural networks. *Computers and Chemical Engineering* **27**, 55–72.

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