Comparative Study of Multicomponent Distillation Static Estimators Based on Industrial and Rigorous Model Datasets

Andrei Torgashov *, Sigurd Skogestad **, Alexey Kozlov ****

*Institute of Automation and Control Process FEB RAS, 5 Radio Str., Vladivostok, Russia (Tel: +7-423-231-02-02; e-mail: torgashov@iacp.dvo.ru).
**Far-Eastern Federal University, Bld. 12(E), Campus FEFU, Russky Island, Vladivostok, Russia.
***Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway (e-mail: skoge@ntnu.no).
****Gazpromneft-Omsk Refinery, JSC, 1, Pr. Gubkina, 644040, Omsk, Russia (e-mail: Kozlov.AV@omsk.gazprom-neft.ru).

Abstract: The comparative study of static estimators (soft sensors) for a multicomponent distillation process based on the industrial and calibrated rigorous model datasets is considered. The sequence of distillation columns of an FCC unit is analyzed as an industrial case study. The contribution of the paper is to develop a method aimed at incorporating a priori knowledge about process in terms of rigorous models for static estimator design when the training sample is small and contains measurement errors. The superiority of a constrained optimization approach for SE design over conventional robust M-estimator is reported. The system of constraints is derived from a calibrated rigorous model of an industrial plant.

Keywords: Soft sensing, estimators, process models, distillation columns.

1. INTRODUCTION

One efficient way to get additional benefits or to reduce the production cost of an industrial company involves the application of an advanced process control (APC) system. The static estimators (SE) or soft sensors (SS) belong to the main functional part in the APC system structure. The estimation of product quality via SE (Ghadrdan et al., 2013) is an inexpensive and attractive technique in industrial automation. The improving of methods for SE, SS evaluations may be a subject of any innovated APC platform (Amrit et al., 2015). In the present work, the sequence of industrial multicomponent distillation columns of the gas separation section of the FCC unit of the Gazpromneft-Omsk Refinery plant is investigated. The widespread approach in industry for SS evaluation is to use statistical methods or a data-driven approach (Fortuna et al., 2007; Kadlec et al., 2009). Most of them are based on the “black-box” concept. In practice, the following obstacles are encountered: small training dataset and measurement (lab) errors. These obstacles were not investigated in the previous papers (Mejdell, T. and Skogestad, 1993; Porru et al., 2013; etc.). Moreover, the unmeasured feed composition disturbances and efficiency variation of separation stages causes problems for obtaining a reliable training sample. Also, some of the key (informative) inputs have low variability ranges due to their seldom varying by operator (for instance, set point of top pressure of the distillation column). This leads to statistical insignificance (t-test) of the corresponding coefficients in the regression model and may not be reconciled with physical-chemical meaning of parameters of SE. The utilization of a calibrated rigorous multicomponent distillation model is helpful to overcome the abovementioned difficulties. It allows estimation of the ranges of parameter values of SE (Torgashov and Zmeu, 2015). The developing approach here is close to the concept of “gray-box” modeling. It has a merit that the rigorous model dataset may give a priori information on the SE equation in terms of admissible parameter ranges in particular for the case of a small industrial training sample. The accuracy of predictions of product quality by SE is investigated by calculation of the coefficient of determination ($R^2$) and root mean squared error (RMSE) for both training and test samples. The indicators $R^2$ and RMSE are widespread in industry and used as main criteria for SE performance.

The paper is organized as follows. The second section contains an industrial process description and statement of the problem. The issues of observability analysis of multicomponent distillation are reported in section 3 from an SE evaluation point of view. The calibration procedure of a rigorous distillation model is presented in section 4. The comparative analysis of SE parameters obtained from industrial and calibrated rigorous model datasets is described in section 5. The constrained optimization approach for SE design is also reported. A description of the method of dataset integration from rigorous modelling into the procedure for SE evaluation based on the raw industrial data is given. In conclusion, the superiority of joint usage of both types of datasets (industrial and from the rigorous model) is summarized.
2. INDUSTRIAL PROCESS DESCRIPTION AND STATEMENT OF PROBLEM

Considered in this paper, the gas separation section of the fluidized catalytic cracking (FCC) unit is represented by two multicomponent distillation columns K-8 and K-9 (Fig. 1). The feed flow comes from the FCC absorption unit and enters on the 25th tray of K-8. The overhead product of K-8 is propane-propylene fraction (PPF). The butane-butylene fraction (BBF) is withdrawn from the top of K-9. The absorption gasoline (AG) is a residue of the gas separation unit and recycled to the absorption section of FCC. The main process variables of the industrial distillation unit are shown in Table 1 and may be considered as informative inputs of multicomponent distillation columns K-8 and K-9 (Fig. 1).

The material balances and feed composition for nominal operating point are presented in Table 2.

Table 1. Main process variables

<table>
<thead>
<tr>
<th>Process variable</th>
<th>Notation</th>
<th>SE variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-8 25th tray pres., kgf/cm²</td>
<td>PIR_7</td>
<td>$x_1$</td>
</tr>
<tr>
<td>K-8 bot. pressure, kgf/cm²</td>
<td>PIR_2</td>
<td>$x_2$</td>
</tr>
<tr>
<td>K-8 bot. temperature, °C</td>
<td>TIRCA_H12</td>
<td>$x_3$</td>
</tr>
<tr>
<td>K-9 bot. pressure, kgf/cm²</td>
<td>PIR_5</td>
<td>$x_4$</td>
</tr>
<tr>
<td>K-9 top temperature, °C</td>
<td>TIRC_7</td>
<td>$x_5$</td>
</tr>
<tr>
<td>K-9 25th tray temp., °C</td>
<td>TIR_13</td>
<td>$x_6$</td>
</tr>
<tr>
<td>Sum of $C_j$ in BBF, %</td>
<td>-</td>
<td>$y$</td>
</tr>
</tbody>
</table>

The mass balance in Table 2 for nominal operating point is derived by averaging industrial data based on the available process statistics. The main goal of the paper is to develop an approach for soft sensor design based on the industrial data and rigorous modeling. It is necessary to overcome such difficulties as small training dataset and laboratory errors. The sum of concentrations of propylene and propane ($C_3$) in BBF is considered as SE output.

3. OBSERVABILITY ANALYSIS OF MULTICOMPONENT DISTILLATION PROCESS

Observability analysis is a key issue for SE and SS design. Traditional observability analysis considers whether all the process states are identifiable. However, in our case, only the top and bottom compositions are interesting states for analysis. We consider a simplified model of the multicomponent distillation process for one column but cover important principles of the industrial unit.

$$\frac{d\tilde{x}_{ji}}{dt} = \frac{(R + D)\alpha_i\tilde{x}_{ji}}{\alpha_c + \sum_{i=1}^{n}\tilde{x}_{ji}(\alpha_i - \alpha_c)} - \tilde{x}_{ji} - D\tilde{x}_{ji}; \quad (1)$$

$$\frac{d\tilde{x}_{ji}}{dt} = R\tilde{x}_{ji} + \frac{(R + D)\alpha_i\tilde{x}_{ji}}{\alpha_c + \sum_{i=1}^{n}\tilde{x}_{ji}(\alpha_i - \alpha_c)} - \tilde{x}_{ji}; \quad (2)$$

$$\frac{d\tilde{x}_{ji}}{dt} = R\tilde{x}_{ji} + \frac{(R + D)\alpha_i\tilde{x}_{ji}}{\alpha_c + \sum_{i=1}^{n}\tilde{x}_{ji}(\alpha_i - \alpha_c)} + F\tilde{x}_{ji}. \quad (3)$$

<table>
<thead>
<tr>
<th>№</th>
<th>Component</th>
<th>Feed, kg/hr</th>
<th>PPF, kg/hr</th>
<th>BBF, kg/hr</th>
<th>AG, kg/hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ethane</td>
<td>23.47</td>
<td>23.47</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>propane</td>
<td>6837</td>
<td>6732</td>
<td>105</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>propane</td>
<td>1629</td>
<td>1342</td>
<td>287</td>
<td>0.0</td>
</tr>
</tbody>
</table>
\[ j = f+1 \ldots K-1: \]
\[ \frac{d\tilde{x}_{ji}}{dt} = (R + F)\tilde{x}_{ji} - \frac{(R + D)\alpha_j\tilde{x}_{ji}}{\alpha_j + \sum_{i=1}^{c-1} \tilde{x}_{ji}(\alpha_i - \alpha_j)}, \]
\[ -(R + F)\tilde{x}_{ji} = \frac{(R + D)\alpha_j\tilde{x}_{ji}}{\alpha_j + \sum_{i=1}^{c-1} \tilde{x}_{ji}(\alpha_i - \alpha_j)}; \]
\[ j = K: \]
\[ \frac{d\tilde{x}_{Kj}}{dt} = (R + F)\tilde{x}_{Kj} - (F - D)\tilde{x}_{Kj} - \frac{(R + D)\alpha_j\tilde{x}_{Kj}}{\alpha_j + \sum_{i=1}^{c-1} \tilde{x}_{Kj}(\alpha_i - \alpha_j)}, \]
\[ \alpha_j + \sum_{i=1}^{c-1} \tilde{x}_{Kj}(\alpha_i - \alpha_j) \]

where \( i = 1 \ldots c; c \) – number of separated compounds; \( K \) – total number of separation stages (trays); \( \tilde{x}_{ji} \) – concentration of \( i \)th compound on the \( j \)th tray in the liquid; \( \tilde{x}_{Fj} \) – concentration of \( i \)th compound in the feed (the feed is entered to the column in the liquid phase); \( R \) – reflux flowrate; \( F \) – feed flowrate; \( D \) – distillate flowrate; \( \alpha \) – relative volatility.

The system of differential equations (1)-(5) may be linearized and represented in the state-space form. The state vector is formulated as \( x = \left( \tilde{x}_{i1} \ldots \tilde{x}_{ij} \ldots \tilde{x}_{i1,1} \ldots \tilde{x}_{i1} \ldots \tilde{x}_{i1} \ldots \tilde{x}_{i1,1} \ldots \right)^T \).

The temperature on each stage \( j \) may be expressed via Antoine equation:
\[ T_j = \frac{B_h + C_h \log \left( \frac{P_j}{\gamma_j} \right) - C_A B_b}{A_b - \log \left( \frac{P_j}{\gamma_j} \right)}, \]
where \( \gamma_j = \alpha + \sum_{i=1}^{c-1} \tilde{x}_{i1}(\alpha_i - \alpha_j); P_j \) – stage pressure; \( A_b, B_b \) and \( C_b \) – constants of Antoine equation for base compound.

The system (1)-(5) reduced to the following form (in the neighborhood of the operating point):
\[ \dot{x} = Ax + Bu + Dd \]
\[ y = Cx, \quad z = Zy \]
\[ h = Hx, \]
where \( z \) is available temperature measurements; \( h \) contains elements of state vector of top and bottom composition; \( Z \) and \( H \) are transforming matrices. The matrix
\[ C = \left( \frac{\partial T_j}{\partial \tilde{x}_{Kj}} \right)_{K(c-1)}^{x_{K(c-1)}} \]
is obtained as
\[ k=x:j: \]
\[ \frac{\partial T_j}{\partial \tilde{x}_{Kj}} = -\frac{(\alpha_i - \alpha_j)(T_j + C_c)}{\gamma_j A_b - \log \left( \frac{P_j}{\gamma_j} \right) ln10}, \]
\[ k=J: \]
\[ \frac{\partial T_j}{\partial \tilde{x}_{Kj}} = 0. \]

Let us consider an illustrative example of obtaining the structure of matrix \( C \) due to its importance for connecting measurements with the state vector. For simplicity, assume that components in the feed can be divided by 4 groups (in terms of boiling temperatures of compounds) with the key individual component in each group and suppose that \( c=4 \) and \( K=5 \):
\[ C = [-60.8 -24.8 -11.5 0 0 0 0 -76.4 -31.1 -14.4 0 0 0 0 0 0 -93.7 0 0 0 0 0 0 0 0 0 0 0 -38.2 -17.7 0 0 0 0 0 0 0 -109.4 -44.6 -20.6 0 0 0 0 0 0 0 0 0 0 0 0 0 -135.2 -55.11 -25.5]. \]

The disturbance vector \( d \) is unknown and associated with the unmeasured feed composition fluctuations. Based on the vectors \( u \) and \( d \), the steady-state solution (for SE) of \( x \) can be obtained as
\[ h = H(-A)^{-1}(Bu + Dd). \]

From equation (6), it follows that the solution of \( h \) is not unique, i.e., under fixed \( <H,A,B,u,D> \), we have infinite numbers of \( h \) for the case of infinite numbers of values of unknown vector \( d \). This means that it is impossible to observe \( h \) via input \( u \).

From the other side, consider the inference of \( h \) via matrix \( C \) and vector of measurements \( y \) (temperatures). The equation for output \( y \) may be written via block-diagonal form of matrix \( C \) as
\[ \begin{pmatrix} T_1 & c_1 & 0 & 0 \\ T_j & 0 & c_j & 0 \\ T_K & 0 & c_K & 0 \end{pmatrix} \]
\[ \begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_j \\ \tilde{c}_K \end{pmatrix} \]
where \( c_j = \left( \frac{\partial T_j}{\partial \tilde{x}_{Kj}} \right) \) - single string matrix;
\[ \tilde{x}_j = \left( \tilde{x}_{j1} \ldots \tilde{x}_{j1} \right)^T \] - vector of liquid composition on stage \( j \).

For the case of top and bottom composition,
\[ \begin{pmatrix} T_1 & c_1 & 0 & 0 \\ T_K & 0 & c_K & 0 \end{pmatrix} \]
\[ \begin{pmatrix} \tilde{c}_1 \\ \tilde{c}_K \end{pmatrix} \]
\[ \begin{pmatrix} \tilde{c}_1 \end{pmatrix} \]

From equation (7), it is obvious that \( h \) cannot be reconstructed (observed) because matrix \( G \) is non-square and has dimension \( 2 \times K(c-1) \). We can find the solution only via
pseudo-inversion of $G$ (to compute $G^+$) in order to get $h$. It is well known that the pseudoinverse provides a least squares solution for (7) (Albert, A., 1972; Lawson, C. and Hanson, R., 1974). This motivates the building of regression models of SE.

4. CALIBRATION OF RIGOROUS DISTILLATION MODEL ON INDUSTRIAL DATA

The motivation of the use of a try-by-tray rigorous (with physical-chemical essence) distillation model implies that it is necessary to extend the training sample by input variables having low variability ranges and investigate the soft sensor model without measurement errors. The main target is to get the uncertainty-free threshold value of identifiability index based on the calibrated rigorous distillation model. This is the principal difference from the entirely statistical approach (e.g., based on the bootstrapping of samples containing errors) of obtaining the threshold value given by Hengl et al. (2007). The principles of mass, energy balances, and phase equilibrium equations of a rigorous model of multicomponent distillation are well known.

The fitting procedure involves the selection of the tray efficiency minimizing the mismatch between industrial data and the rigorous model. When estimating Murphree efficiency on industrial data, the average steady-state operating point is involved.

The following residual function is utilized: $f_{ij}(E1, E2) = (x_{ij}(E1, E2) - x_{ij}^{ind})^2$, where $x_{ij}^{ind}$ is isopentane content calculated using rigorous model; $x_{ij}$ is isopentane industrial data; $E1$ – absorption section efficiency; $E2$ – stripping section efficiency. Moreover, the data reconciliation problem of reflux ratio (RR) is solved in conjunction with the residual function minimization.

The calculation results of fitting procedure and Murphree efficiencies’ feasible domain are shown in Fig. 3. The point * with coordinates of $E1 = 0.608$ and $E2 = 0.5$ indicates the best fitted values of sectional efficiencies with optimal reconciled RR value. It will be noted that the fitting of efficiency values for each individual separation stages did not provide significant improvements of calibration results.

5. COMPARATIVE STUDY OF STATIC ESTIMATORS

The SE equation is considered in the following linear form:

$$\hat{y} = \hat{b}_0 + \sum_{i=1}^{m} \hat{b}_i x_i,$$

where $\hat{y}$ – SE output (prediction); $x_i$ - measured input of SE (Table 1); $\hat{b}_i$ - SE model parameter; $m$ – number of inputs.

The comparative analysis of model parameters of SE for $C_3$ prediction in BBF may be done from Table 4. It is interesting to note that the $1^{st}$ and $4^{th}$ parameters ($b_1$ and $b_4$) have different signs but are still statistically significant (t-test) for both models.

![Fig. 2. The results of fitting the sectional efficiencies on industrial data.](image-url)

Based on the fitted rigorous model of industrial distillation unit, it will be possible to generate a training dataset. But the main problem for multicomponent mixture was to generate the feed composition vectors for simulation, which are close to industrial situations. For that reason, we study the redistribution of individual components among the main separated groups C3 and C4 in the feed composition (Table 3 and Figure 3). It was helpful to reduce the numerical experiments with a calibrated rigorous model. For high correlated concentrations of components in the feed (e.g., propylene and propane), it is possible to express variation of both compounds via one.

![Fig. 3. Redistribution of individual components in the feed composition among the main groups C3/C4.](image-url)

### Table 3. Correlation coefficients of feed components in the groups C3 and C4

<table>
<thead>
<tr>
<th></th>
<th>propylene</th>
<th>propane</th>
<th>t-butane</th>
<th>n-butane</th>
<th>t-butene-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>propane</td>
<td>0.967</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-butane</td>
<td>0.096</td>
<td>0.106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-butane</td>
<td>0.067</td>
<td>0.155</td>
<td>0.803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-butene-2</td>
<td>-0.006</td>
<td>0.036</td>
<td>0.703</td>
<td>0.817</td>
<td></td>
</tr>
<tr>
<td>c-butene-2</td>
<td>0.040</td>
<td>0.087</td>
<td>0.066</td>
<td>0.782</td>
<td>0.833</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>propylene</th>
<th>propane</th>
<th>t-butane</th>
<th>n-butane</th>
<th>t-butene-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>propane</td>
<td>0.967</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n-butane</td>
<td>0.096</td>
<td>0.106</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-butane</td>
<td>0.067</td>
<td>0.155</td>
<td>0.803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-butene-2</td>
<td>-0.006</td>
<td>0.036</td>
<td>0.703</td>
<td>0.817</td>
<td></td>
</tr>
<tr>
<td>c-butene-2</td>
<td>0.040</td>
<td>0.087</td>
<td>0.066</td>
<td>0.782</td>
<td>0.833</td>
</tr>
</tbody>
</table>

### Table 4. Comparison of model parameters

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Model 1 (based on industrial data): $b_{ind}$</th>
<th>Model 2(based on rigor.simul. data): $b_{rig}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>9.98</td>
<td>-0.06</td>
</tr>
<tr>
<td>$b_2$</td>
<td>2.67</td>
<td>2.63</td>
</tr>
</tbody>
</table>
Both models form Table 4 are derived by use of robust regression $M$-estimator (Maronna et al., 2006) without constraints on the parameter values. The main advantage of model 2 is the absence of influence of laboratory errors and incorrect input values (due to sampling time uncertainty) on the parameters of SE.

The least-squares fitting procedure under restriction of model parameters is transformed into optimization problem

$$
\min_{\mathbf{b}} \frac{1}{2} \| \mathbf{Xb} - \mathbf{y} \|^2 \quad \text{such that} \quad \mathbf{l}_b \leq \mathbf{b} \leq \mathbf{u}_b,
$$

where $\mathbf{b} = (b_1 \ldots b_m)^T$ – vector of model parameters; $\mathbf{X}$ – input data matrix; $\mathbf{y}$ – vector of output measurements; $\mathbf{l}_b$ and $\mathbf{u}_b$ are lower and upper bounds of SE parameter values.

Figures 4 and 5 contain $R^2$ comparison analysis of $M$-estimators for a conventional unconstrained (robust reg.) case and constrained optimization approach. The vectors of the lower and upper bounds of parameters of model 2 ($\mathbf{b}_{rig}$) from Table 4 are derived as follows:

$$
\mathbf{l}_b = 0.7 \times \mathbf{b}_{rig} \quad \text{and} \quad \mathbf{u}_b = 1.3 \times \mathbf{b}_{rig},
$$

where the 30% variation range is assigned for vector $\mathbf{b}_{rig}$ from industrial experience point of view.

In Figures 6 and 7, the results of the $RMSE$ study are depicted. The investigation of industrial data is organized as follows for Figs. 4-7. The total sample of data contains 179 observations. The initial sample has 70 observations and is divided into training (70%) and test (30%) datasets. Starting from 70 and up to 179, each sample is separated into a 70/30 percent ratio in order to get the training and test datasets. The SE model evaluation procedure (with $R^2$ and $RMSE$ calculations) is repeated for each dataset and actually represents the adaptive performance of SE.

The adjustment of selected parameters ($b_1$ and $b_4$) of SE (using unconstrained $M$-estimator and constrained optimization) is plotted in Figs. 8-9. Obviously, the inclusion of constraints of estimated parameters provides the better performance of SE for testing datasets. Table 5 summarizes calculation with industrial data.
The improvements of SE performance for testing datasets of \( R^2 \) and \( \text{RMSE} \) are \( 100\times(0.81-0.36)/0.81=55\% \) and \( 100\times(1.17-0.32)/1.17=72\% \), respectively.

### 6. CONCLUSIONS

The article addressed the problem of improving SE evaluation. The way to take into account all available a priori information about the physics-chemical meaning in the frame of rigorous model has been proposed. Finally, it gives the possibility of an increase of \( R^2 \) on the test sample.

The calibrated rigorous multicomponent distillation model is useful for checking the correctness of SE parameter values and to obtain \( b_{\text{rig}} \). The introduction of vectors \( l_b \) and \( u_b \) may be considered as an indirect gray-box modelling approach for SE evaluation for multicomponent distillation columns. The selection of inputs of SE also becomes reliable in that case and reconciled with thermodynamic essence. As a result, the improvement of SE quality leads to successful industrial application of APC systems.

### 7. ACKNOWLEDGEMENTS

This work is partially supported by the Far Eastern Branch of the Russian Academy of Sciences under the “Far East” program (project No. 0262-2015-0063).

### REFERENCES


