

# A Rigorous Model for Evaluating Moving Window Soft Sensors for Industrial Distillation Processes

Andrei Torgashov<sup>a,\*</sup>, Sigurd Skogestad<sup>b</sup>, Dmitry An<sup>c</sup>

<sup>a</sup>Process Control Laboratory, Institute of Automation and Control Process FEB RAS, 5 Radio Str., Vladivostok, Russia

<sup>b</sup>Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, NO-7491, Norway

<sup>c</sup>Gazpromneft-Omsk Refinery, JSC, 1, Pr. Gubkina, 644040, Omsk, Russia

torgashov@iacp.dvo.ru

Traditionally, soft sensors are developed based on measurement data only, but here we consider a moving window (MW) soft sensor (SS) that uses data generated from a calibrated, rigorous model of the distillation columns of an FFC unit at Gazpromneft-Omsk Refinery. The contribution of the paper is that a procedure is developed for MW SS design that incorporates a priori knowledge, which is especially suitable when the training sample is small and contains measurement errors. In addition, we propose a continuous adaptation of all model parameters based on new data, instead of the usual procedure of only updating the bias. The accuracy of the predicted product quality is investigated by calculating the coefficient of determination ( $R^2$ ) and root mean squared error (RMSE) for the test sample. Several approaches were considered, and we found that a constrained optimization approach was superior. The constraints on the model parameters of SSs are derived from a calibrated, rigorous distillation unit model. The improved estimator quality resulted in the successful industrial application of advanced process control (APC) systems.

## 1. Introduction

One efficient way for an industrial company to gain additional benefits or reduce production costs is to apply an APC system. Soft sensors (SSs) belong to the main functional part of the APC system's structure. The estimation of product quality via SSs (Ghadrdan et al., 2013) is an inexpensive and attractive technique in industrial automation. Improving SS evaluation methods may be the subject of any innovated APC or real-time optimization platform (Amrit et al., 2015; Fayruzov et al., 2017; Wang et al., 2017).

In the present work, the sequence of industrial multicomponent distillation columns in the gas separation section of the fluidized catalytic cracking (FCC) unit at 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; Kadlec et al., 2011). Most of them are based on the "black-box" concept. In practice, several obstacles are encountered, including a small training dataset, measurement errors, and an unsteady mode of distillation operation. Moreover, unmeasured feed-composition disturbances and variations in the efficiency of separation stages cause problems for obtaining a reliable training sample. Also, some of the key (informative) inputs have low variability ranges due to the operator seldom varying them (for instance, the set point for the top pressure of the distillation column). This leads to statistical insignificance (via t-test) of the corresponding coefficients in the regression model and may not be reconciled with the physical-chemical meaning of the parameters of SSs. A calibrated rigorous multicomponent distillation model is helpful to overcome these difficulties. It allows the ranges of the SSs' parameter values to be estimated (Torgashov and Zmeu, 2015).

The developed approach here is close to the concept of "gray-box" modeling. One of its merits is that the rigorous model dataset may give a priori information on the SS equation in terms of admissible parameter ranges, particularly for a small industrial training sample. The accuracy of product quality predictions by SSs is investigated by calculating the coefficient of determination ( $R^2$ ) and root mean squared error (RMSE) for the test sample.

The paper is organized as follows. The second section describes an industrial process and states the problem. The calibration procedure for a rigorous distillation model is presented in section 3. A comparative analysis of the SS parameters obtained from industrial and calibrated rigorous model datasets is described in section 4. The constrained optimization approach for SS design is reported in section 5. A description of how the dataset was integrated from rigorous modelling into the SS evaluation procedures based on the raw industrial data is given. In conclusion, the superiority of using both industrial and rigorous-model datasets for MW SS design is summarized.

## 2. Industrial process description and statement of the problem

The gas-separation section of the FCC unit considered in this paper 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 25<sup>th</sup> 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. Absorption gasoline (AG) is a residue of the gas-separation unit and is recycled in the absorption section of FCC unit. The main process variables of the industrial distillation unit are shown in Table 1 and may be considered as informative inputs of the SS.

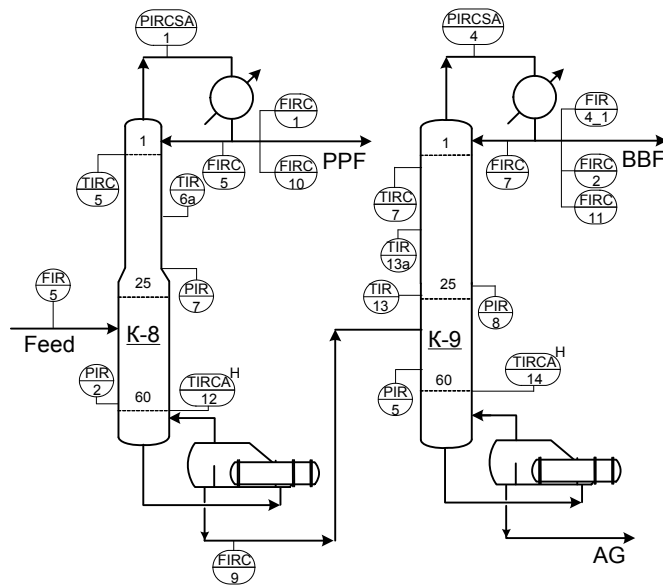


Figure 1: Sequence of industrial multicomponent distillation columns

Table 1: Main process variables

Process variable	Notation	SS variable
K-8 25 <sup>th</sup> tray pres., kgf/cm <sup>2</sup>	PIR_7	x <sub>1</sub>
K-8 bot. pres., kgf/cm <sup>2</sup>	PIR_2	x <sub>2</sub>
K-8 bot. temp., °C	TIRCAH_12	x <sub>3</sub>
K-9 bot. pres., kgf/cm <sup>2</sup>	PIR_5	x <sub>4</sub>
K-9 top temp., °C	TIRC_7	x <sub>5</sub>
K-9 25 <sup>th</sup> tray temp., °C	TIR_13	x <sub>6</sub>
Sum of C <sub>3</sub> in BBF, %	-	y

The material balances and feed composition for the nominal steady-state operating point are presented in Table 2. The mass balance in Table 2 for the nominal operating point is derived by averaging industrial data based on the available process statistics.

The main goal of this paper is to develop an approach to MW soft-sensor design based on industrial data and rigorous modeling. It is necessary to overcome difficulties such as small training datasets and laboratory errors. The total concentrations of propylene and propane (C<sub>3</sub>) in the BBF is considered the SS output.

Table 2: Feed composition and material balances of the multicomponent distillation unit

No.	Component	Feed, kg/hr	PPF, kg/hr	BBF, kg/hr	AG, kg/hr
1	ethane	23.47	23.47	0.0	0.0
2	propylene	6837	6732	105	0.0
3	propane	1629	1342	287	0.0
4	i-butane	4610	233	4377	0.0
5	butene-1	5051	50	5001	0.0
6	n-butane	1043	0.0	1042	1.0
7	t-butene-2	2611	0.0	2604	7.9
8	c-butene-2	1876	0.0	1840	36
9	3-methylbutene-1	95.9	0.0	0.0	95.9
10	i-pentane	571	0.0	0.0	571
11	2-methylbutene-1	182	0.0	0.0	182
12	n-pentane	24.2	0.0	0.0	24.2
13	t-pentene-2	71.8	0.0	0.0	71.8
14	c-pentene-2	80.4	0.0	0.0	80.4
15	3,3-dimeth.but.-1	45.1	0.0	0.0	45.1
16	2-methylpentane	5.54	0.0	0.0	5.54
17	3-methylpentane	1.97	0.0	0.0	1.97
18	n-hexane	2.70	0.0	0.0	2.70
19	benzene	0.89	0.0	0.0	0.89
20	2-methylhexane	0.46	0.0	0.0	0.46

### 3. Calibration of rigorous distillation model on industrial data

A rigorous tray-by-tray (with physical-chemical essence) distillation model was used because it is necessary to extend the training sample with input variables that have low variability ranges and investigate the soft sensor model without measurement errors.

The principles of mass, energy balances, and phase-equilibrium equations of a rigorous model of multicomponent distillation are well known. The fitting procedure involves selecting the value of tray efficiency that minimizes the mismatch between industrial data and the rigorous model. The average (nominal) steady-state operating point is involved when estimating the Murphree efficiency with industrial data. The following residual function is utilized:

$$f_{ic5}(E1, E2) = (x_{ic5}^{rm}(E1, E2) - x_{ic5})^2,$$

where  $x_{ic5}^{rm}(E1, E2)$  is the isopentane content calculated using the rigorous model,  $x_{ic5}$  is the isopentane industrial data, E1 is the absorption section's efficiency, and E2 is the stripping section's efficiency. Moreover, the data-reconciliation problem of reflux ratio (RR) is solved in conjunction with the residual function minimization. The calculated results of the fitting procedure and the Murphree efficiency's feasible domain were investigated. The optimal point with coordinates of E1 = 0.608 and E2 = 0.5 indicates the best-fitted values of sectional efficiencies with the reconciled RR value. Notably, fitting efficiency values for each individual separation stage did not significantly improve the calibration results.

Based on the fitted rigorous model of the industrial distillation unit, it will be possible to generate a training dataset. However, the main problem for the multicomponent mixture was generating the feed-composition vectors for simulations that are close to industrial situations. For that reason, we studied the redistribution of individual components among the main separated groups, C<sub>3</sub> and C<sub>4</sub>, in the feed composition. It was helpful to reduce the numerical experiments with a rigorous calibrated model. For highly correlated concentrations of components in the feed (e.g., propylene and propane), it is possible to express the variation of both compounds via one concentration.

### 4. Development of soft sensors based on industrial and rigorous model datasets

The SS model equation is considered in the following linear form:

$$\hat{y} = b_0 + \sum_{k=1}^m b_k x_k,$$

where  $\hat{y}$  is the SS's output (prediction),  $x_k$  is the measured input of the SS (Table 1);  $b_k$  is the SS model parameter, and  $m$  is the number of inputs.

The comparative analysis of the SS model parameters for predicting  $C_3$  concentration in the BBF may be done from Table 3.

Table 3. Comparison of model parameters

Model parameter	Model 1 (based on industrial data): $\mathbf{b}_{ind}$	Model 2 (based on rigorous simulation data): $\mathbf{b}_{rig}$
$b_1$	9.98	-0.06
$b_2$	2.67	2.63
$b_3$	-0.66	-0.97
$b_4$	1.06	-3.82
$b_5$	-1.04	-1.02
$b_6$	1.0	1.53
$b_0$ (bias)	-154.55	47.6

Both models in Table 3 were derived by use of the robust regression  $M$ -estimator (Maronna et al., 2006) without constraints on the parameter values. The industrial dataset was used to evaluate model 1. The training sample for model 2 was generated based on simulations with a calibrated, rigorous model of the industrial distillation unit (section 3). 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 SS parameters. Interestingly, the 1<sup>st</sup> and 4<sup>th</sup> parameters ( $b_1$  and  $b_4$ ) have different signs but were still statistically significant (t-test) for both models. However, the signs and values of  $b_1$  and  $b_4$  are only consistent with the physical-chemical essence of distillation for model 2.

## 5. Evaluation of constrained moving window static estimators

The least-squares fitting procedure under restriction of the model parameters was transformed into an optimization problem:

$$\min_{\mathbf{b}} \frac{1}{2} \|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2^2 \quad \text{such that } \mathbf{l}_b \leq \mathbf{b} \leq \mathbf{u}_b, \quad (1)$$

where  $\mathbf{b} = (b_1 \dots b_m)^T$  is the vector of the model's parameters,  $\mathbf{X}$  is the input data matrix,  $\mathbf{y}$  is the vector of output measurements, and  $\mathbf{l}_b$  and  $\mathbf{u}_b$  are the lower and upper bounds of the SS parameter values, respectively. Vectors  $\mathbf{l}_b$  and  $\mathbf{u}_b$  play an important role in integrating a calibrated first-principle (rigorous) model into the SS design procedure. The values of these vectors may only be obtained numerically because of the high dimensionality of the multicomponent distillation model. The comparative summary of the application of SS for the industrial distillation process can be seen in Tables 4 and 5. Table 4 compares the  $RMSE$  and  $R^2$  values of the  $M$ -estimators for a conventional unconstrained (robust reg.) case and constrained optimization approach. The vectors of the lower and upper bounds of model 2's ( $\mathbf{b}_{rig}$ ) parameters from Table 3 are derived as follows:  $\mathbf{l}_b = (1 - \nu) \cdot \mathbf{b}_{rig}$  and  $\mathbf{u}_b = (1 + \nu) \cdot \mathbf{b}_{rig}$ , where the value of  $\nu$  (variation range parameter) is assigned for vector  $\mathbf{b}_{rig}$  from the solution the of optimization problem ( $RMSE$  is criterion), as given below.

Table 4. MW soft-sensor performance

	Robust reg. with all prev. data	Robust reg. MW	Constrained MW
$R^2$	0.65	0.68	0.81
$RMSE$	0.62	0.60	0.45

Table 5. Variance (std.) of the soft sensors' parameters

Model parameter	Robust reg. with all prev. data	Robust reg. MW	Constrained MW
$b_1$	2.56	4.61	0.02
$b_2$	1.74	3.12	0.48
$b_3$	0.05	0.16	0.09
$b_4$	1.11	5.31	0.63
$b_5$	0.11	0.29	0.30
$b_6$	0.06	0.24	0.15
$b_0$ (bias)	40.63	54.22	19.45

For the testing dataset of  $R^2$  and  $RMSE$ , SS performance improved by  $100 \cdot (0.81 - 0.68) / 0.81 = 16\%$  and  $100 \cdot (0.60 - 0.45) / 0.60 = 25\%$ , respectively.

The variability of the MW soft-sensor parameters was lower in the case of constrained optimization (Figures 2 and 3) for the test dataset. Therefore, the use of vectors  $\mathbf{I}_b$  and  $\mathbf{u}_b$  made the performance of the MW soft sensor more stable and reliable.

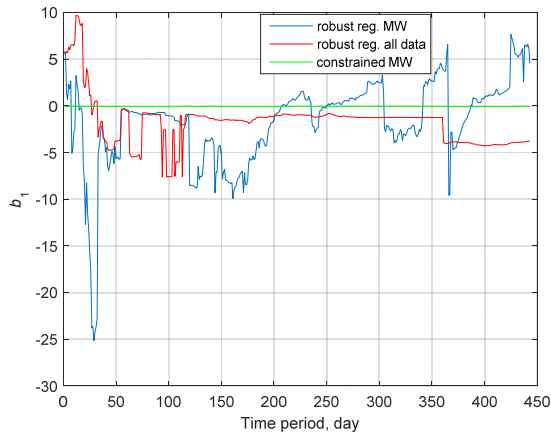


Figure 2: Adaptation of parameter  $b_1$  (test dataset)

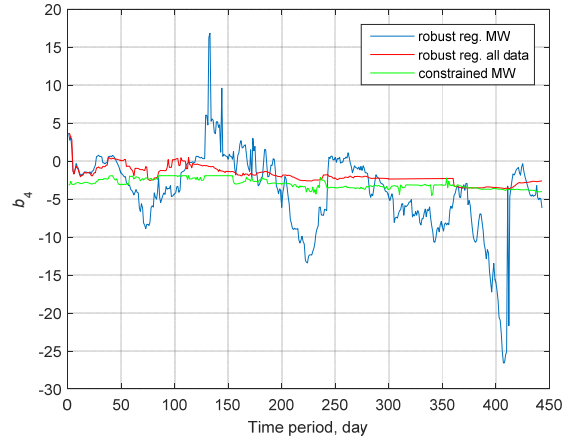


Figure 3: Adaptation of parameter  $b_4$  (test dataset)

The width of MW ( $w$ ) is an important parameter of adaptive soft sensors. The calculation of  $w$  is proposed by the following formula:

$$w = q \cdot m, \quad (2)$$

where  $q$  is the window width parameter.

The optimal parameters of the MW soft sensor (based on the constrained optimization) are shown in Figure 4. The optimal values (in the sense of the  $RMSE$  criterion) of the parameters are  $v^* = 50\%$  and  $q^* = 10$  in (2) for the considered adaptive soft sensor.

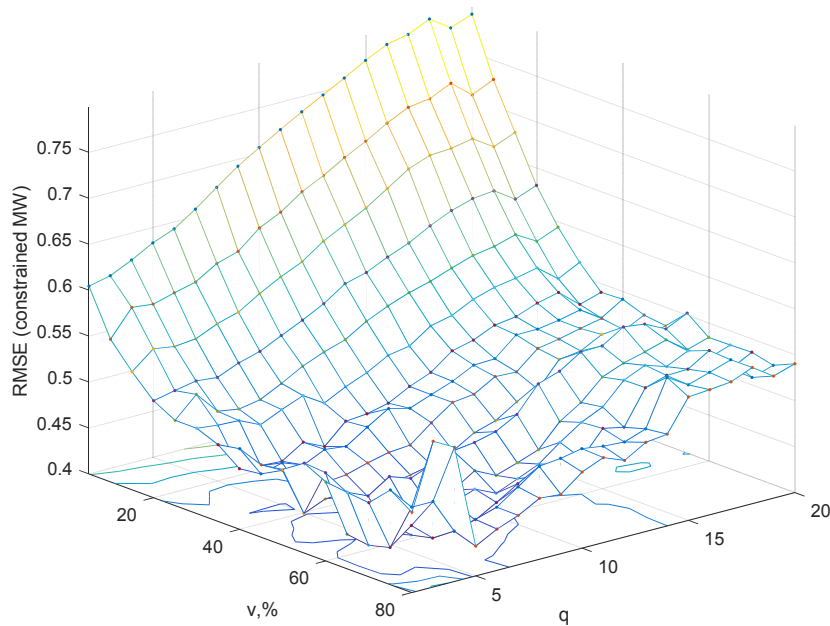


Figure 4: Selection of optimal values of  $q$  and  $v$  for constrained MW SS

The evaluation of a constrained MW SS can be described by the following stepwise procedure:

- Step 1. Divide the industrial data into a training dataset and a test dataset.
- Step 2. Evaluate model 1 based on the training dataset and obtain the vector of parameters  $\mathbf{b}_{\text{ind}}$  (Table 3).
- Step 3. Calibrate (jointly with data reconciliation) the rigorous distillation unit model with the training dataset, as described in section 3.
- Step 4. Evaluate model 2 based on the dataset from the simulator with the calibrated, rigorous distillation unit model and obtain the vector of parameters  $\mathbf{b}_{\text{rig}}$  (Table 3).
- Step 5. Calculate bounds  $\mathbf{l}_b$  and  $\mathbf{u}_b$  based on the vector  $\mathbf{b}_{\text{rig}}$  and study parameter  $v$  belonging to the interval 0.3-0.5. The use of  $\mathbf{b}_{\text{ind}}$  is not valid because the signs and values of  $b_1$  and  $b_4$  (in model 1) are consistent with the physical-chemical essence of distillation.
- Step 6. Analyze the performance of the adaptive soft sensor with the test dataset. Calculate the prediction error (PE) for each new point of the test dataset to solve the optimization problem (1) based on the previous historical data,  $\mathbf{X}, \mathbf{y}$ . The mean value of the set of PEs is the *RMSE* for the test dataset.

For each point of the test dataset, vector  $\mathbf{b}$  is updated using (1). The application of a rigorous distillation unit model is realized via bounds  $\mathbf{l}_b$  and  $\mathbf{u}_b$ .

## 6. Conclusions

This article improves SS evaluation based on the MW adaptation technique. A way to account for all available a priori information about a multicomponent distillation process, along with the physics-chemical meaning in the frame of a rigorous model, was proposed. This method can reduce the *RMSE* of the test sample in conditions when the training sample is small and contains measurement errors.

The calibrated, rigorous multicomponent distillation model is useful for checking the correctness of the SS's parameter values and for obtaining  $\mathbf{b}_{\text{rig}}$ . The introduction of vectors  $\mathbf{l}_b$  and  $\mathbf{u}_b$  may be considered an indirect gray-box modeling approach for evaluating SSs in multicomponent distillation columns. The selection of SS inputs also becomes reliable in that case and is reconciled with the thermodynamic essence of the distillation. Finally, the use of constraints  $\mathbf{l}_b$  and  $\mathbf{u}_b$  reduces the variability of the MW soft-sensor parameters and leads to more stable and accurate estimations of product quality for industrial distillation processes.

## Acknowledgments

This work was partially supported by the Russian Foundation for Basic Research (Project No. 17-07-00235 A).

## References

- Amrit R., Canney W., Carrette P., Linn R., Martinez A., Singh A., Skrovanek T., Valiquette J., Williamson J., Zhou J., Cott B.J., 2015, Platform for Advanced Control and Estimation (PACE): Shell's and Yokogawa's Next Generation Advanced Process Control Technology. Preprints of the 9th International Symposium on Advanced Control of Chemical Processes, The International Federation of Automatic Control, June 7-10, Whistler, British Columbia, Canada, 1–5.
- Fayruzov D.Kh., Bel'kov Yu.N., Kneller D.V., Torgashov A.Yu., 2017, Advanced process control system for a crude distillation unit. A case study, Automation and remote control, Vol 78, No 2, 357–367.
- Fortuna L., Graziani S., Rizzo A., Xibilia M.G., 2007, Soft sensors for monitoring and control of industrial processes, London, Springer-Verlag.
- Ghadrdan M., Halvorsen Ivar J., Skogestad S., 2013, Dynamic compensation of static estimators from Loss method. Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems (DYCOPS), Mumbai, India, Dec. 2013, International Federation of Automatic Control, IFAC-PapersOnLine, ISSN 1474-6670, 403–408.
- Kadlec P., Gabrys B., Strandt S., 2009, Data-driven soft sensors in the process industry, Computers and Chemical Engineering, Vol 33, 795–814.
- Kadlec P., Grbic R., Gabrys B., 2011, Review of adaptation mechanisms for data-driven soft sensors. Computers and Chemical Engineering, Vol 35, 1–24.
- Maronna R. A., Martin R. D., Yohai V. J., 2006, Robust Statistics: Theory and Methods, New York, John Wiley & Sons.
- Torgashov A., Zmeu K., 2015, Nonparametric soft sensor evaluation for industrial distillation plant, Computer Aided Chemical Engineering, Vol 37, 1487–1492.
- Wang X., Mahalec V., Li Z., Qian F., 2017, Real-time optimization and control of an industrial ethylbenzene dehydrogenation process, Chemical Engineering Transactions, Vol 61, 331–336.