Nonparametric Soft Sensor Development For Distillation-In-Series Unit

Igor Mozharovskii[1,2]
[1]Institute of Automation and Control Processes FEB RAS
[2]Vladivostok State University
Vladivostok, Russia
studvvsu@gmail.com

Svetlana Shevlyagina[1,2]
[1]Institute of Automation and Control Processes FEB RAS
[2]Far Eastern Federal University
Vladivostok, Russia
samotylova@dvo.ru

2024 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM) | 979-8-3503-9501-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIEAM60818.2024.10554054

Abstract—**Oil refining is an energy-intensive process that continuously separates petroleum products. Refining companies are interested in products that meet all regulations and the efficiency of their production. It is possible to achieve cost minimization and increase the quality of the output product by using accurate soft sensors that can reliably predict the quality of the output product in real time. In this regard, research on the construction of reliable soft sensors in industrial production is necessary and relevant. To improve the estimation capability, the alternating conditional expectation algorithm, which is based on nonparametric optimal transformations, was used for nonparametric soft sensor design. As a result, the designed nonparametric soft sensor shows a better efficiency in predicting the quality index of the target distillation product with a significantly reduced root mean square error compared to multiple linear regression analysis.**

Keywords—alternating conditional expectation algorithm, nonparametric soft sensor, distillation-in-series unit

I. INTRODUCTION

Due to the increasing demands on the quality of main types of petroleum products, oil refining and petrochemical enterprises must continuously improve the economic efficiency of production and the quality of output products. Production efficiency can be improved by using systems for virtual monitoring and control of output product quality indicators. This requires the development of more accurate soft sensors (SS) for estimating the quality of output products, describing nonlinear processes occurring in industrial plants [1, 2].

Laboratory results and industrial data (data from embedded measurement devices such as thermocouples, pressure sensors, flow rate sensors, etc.) are used to develop soft sensors [3-5]. The use of soft sensors allows engineers to assess the quality of the output product, respond to technical failures in a timely manner and quickly adjust the plant's operating modes in an optimal manner to achieve the best possible productivity and reduce production costs. An additional benefit of soft sensors is the early detection and elimination of substandard products in the production process, which eliminates the need for costly downstream processing of rejected raw materials. Researchers face a number of challenges in developing soft sensors for evaluating quality indicators of output product for industrial distillation columns [6, 7]. These problems are caused by constant changes in the composition of the raw materials fed to the columns, as well as nonlinearity of the process, small size of training sample, poor formalization of the object, etc.

Parametric and non-parametric approaches both are used to build soft sensors. The former methods include regression methods such as least squares [8, 9], robust regression [10-12], ridge regression [13-16], projection regression method [17-18], etc. When using the above methods, it is necessary to assume functional dependence with successive refinement of the values of its coefficients. In the case of nonlinearity of the technological process, the most promising are nonparametric approaches [19-21], in particular, fuzzy logic [22-24], artificial neural networks (ANN) [25-27] and the algorithm of alternating conditional expectations (ACE) [28].

As a rule, soft sensors based on of fuzzy algorithms are often successfully suitable for building models of complex objects, where the input data could be imprecise and weakly formalized. The main disadvantages of this approach are the complexity and subjectivity of the formation of fuzzy logic rules, which form the basis of the model functioning, as well as the choice of the type and parameters of its membership functions. This requires empirical knowledge about the modeled object and a big training sample of data.

The approach of building soft sensors based on neural networks has proven to be effective for complex industrial units. However, further research is needed to determine the structure and parameters of the prediction algorithm of neural networks, on which the success of the task solution depends [29].

The algorithm of conditional alternating mathematical expectations allows to find complex dependencies of input variables on output variables, to determine the structure of relationships whose type is initially unknown, and to identify nonlinear functional dependencies based on transformations of the variables used [30, 31].

In this work, the ACE algorithm is used for distillation-in-series unit being a weakly formalized object to improve the SS estimation capability.

The research was carried out within the state assignment of IACP FEB RAS (Theme FWFW-2022-0002).

Therefore, the purpose of this paper is to use nonlinear approach to model the relationship between manipulated variables and finally to predict the quality of the output product with high accuracy. The paper is organized as follows. In Sec. II, we briefly describe the distillation-in-series unit we use. Sec. III presents the nonlinear regression approach we use. Sec. IV contains the application to the real industrial unit and finally in Sec. V we discuss our results.

II. Description of the Naphtha Stabilization Process

The distillation-in-series unit consists of three consecutive distillation columns (Fig. 1) and is designed for naphtha stabilization with production of hydrocarbon gas, fraction 35°C, fraction 35-70°C and fraction 70-140°C. Unstable naphtha enters the 22nd tray of column 1 (C-1). Vapors from the top of C-1 are condensed and collected in the reflux drum, where the hydrocarbon gas exits. The stream of top product (distillate) is withdrawn. The liquid from the bottom tray enters the 29th tray of column 2 (C-2), where separation into the 70°C and 70-140°C fractions takes place. The 70°C fractions enter the 41st tray of column 3 (C-3), where the 35°C and 35-70°C fractions are separated.

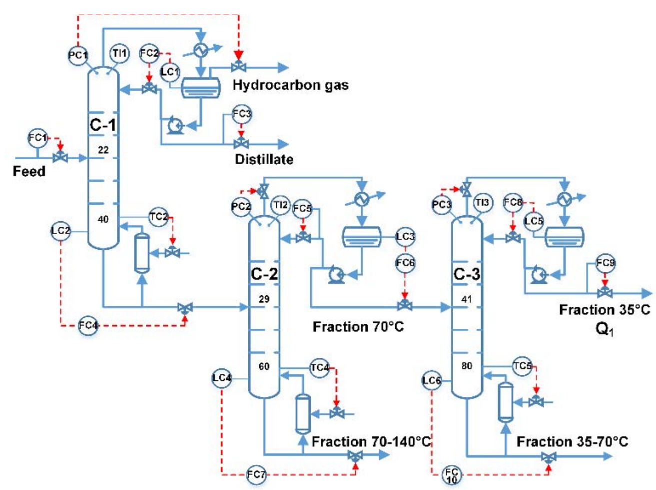

Fig. 1. Schematic of the naphtha stabilization process.

The quality of the 35-70°C fraction is determined by the sum of the C_1-C_4 contents. The details of input and output process variables are presented in Table 1.

TABLE I. Process Variables for the distillation-in-series Unit

No.	Description	Tag	Variable
1	The top pressure of the Column 1 (kgf/cm²)	PC-1	x_1
2	The distillate flow from Column 1 (kg/h)	FC-3	x_2
3	The top pressure of the Column 3 (kgf/cm²)	PC-3	x_3
4	The bottom temperature of the Column 3 (°C)	TC-5	x_4
5	The ratio of the reflux flow to Column 3 (kg/h) to the feed flow to Column 3 (kg/h)	$\frac{FC-8}{FC-6}$	x_5
6	sum C_1-C_4 contents in the bottom stream Column 3 (mass.-%)	Q_1	y

For inputs variables x_i, $i = \overline{1,k}$, and output (response) y, soft sensor should be obtained as functional dependence. Assume a nonlinear model of the form:

$$y = f(x_1, \cdots, x_k) + \varepsilon, \tag{1}$$

where x_i and y are one dimensional vectors, ε is random error component.

III. The Alternating Conditional Expectations Algorithm

To improve estimation capability for weakly formalized objects the ACE algorithm was chosen. For $x_1, \ldots, x_k$ that are independent variables (predictors) and y being a response variable, the ACE regression model [32, 33] can be written as:

$$\theta(y) = \sum_{i=1}^{k} \phi_i(x_i), \tag{2}$$

where θ is a function of the response variable y, ϕ_i are functions of the predictors x_i, $i = \overline{1,5}$.

If θ is invertible, the estimated model ($\hat{y}$) in (2) can be written as:

$$\hat{y} = \hat{\theta}^{-1} \sum_{i=1}^{k} \hat{\phi}_i(x_i) \tag{3}$$

The ACE algorithm minimizes the squared error:

$$e^2(\theta, \phi_1, \ldots, \phi_k) = E\left[\theta(y) - \sum_{i=1}^{k} \phi_i(x_i)\right]^2 \tag{4}$$

Keeping $E\theta^2 = 1$, $E\theta = E\phi_1 = \ldots = E\phi_p = 0$ the minimization of e^2 is carried out through a series of single-function minimizations:

$$\theta(y) = \frac{E\left[\sum_{i=1}^{k} \phi_i(x_i)|y\right]}{\left\|E\left[\sum_{i=1}^{k} \phi_i(x_i)|y\right]\right\|}, \tag{5}$$

$$\phi_i(x_i) = E\left[\theta(y) - \sum_{j \neq i}^{k} \phi_j(x_j)\, |x_i\right]. \tag{6}$$

The final $\phi_i(x_i)$ and $\theta(y)$ after the minimization are estimates of the optimal transformations $\hat{\phi}_i(x_i)$ and $\hat{\theta}(y)$. In the transformed space, the response and predictor variables are related as follows:

$$\hat{\theta}(y) = \sum_{i=1}^{k} \hat{\phi}_i(x_i). \tag{7}$$

The iteration process to find the optimal transformations begins with an initial assignment for one of the functions ($\theta = y/\|y\|$, where the norm is defined by $\|y\| = \sqrt{var(y)}$) and ends when the full iteration cycle no longer leads to a decrease in e^2 (4). Pseudo-code of the ACE algorithm is presented below.

Algorithm 1. Pseudo-code of the ACE algorithm.

Start

input: $\theta(y) = y/\|y\|$ and $\phi_1(x_1), \ldots, \phi_k(x_k) = 0$;

while $e^2(\theta, \phi_1, \ldots, \phi_p) \to min$

while $e^2(\theta, \phi_1, \ldots, \phi_p) \to min$

for $p = 1{:}\,k$

$\hat{\phi}_p(x_i) = E\left[\theta(y) - \sum_{j\neq i}^{k} \phi_j(x_j) | x_i\right]$;

$\phi_i(x_i) = \hat{\phi}_i(x_i)$;

end

end

$\hat{\theta}(y) = E\left[\sum_{i=1}^{k} \phi_i(x_i) | y\right] / \left\| E\left[\sum_{i=1}^{k} \phi_i(x_i) | y\right]\right\|$

$\theta(x) = \hat{\theta}(y)$

end

output: the final $\theta(y)$, $\phi_i(x_i)$ are estimates of the optimal transformations $\hat{\theta}(y)$, $\hat{\phi}_i(x_i)$.

End

These transformations are estimated by using a smoothing technique for example different kernel estimators. We use a modified version of ACE as shown in [20], where the data are ranked before optimal transformations, which are estimated (we sort the data set $x_1, \ldots, x_k$ in ascending order, resulting in the vector y, and all further calculations are performed with the corresponding index vector I, where $y = x_i(I)$). This allows a more accurate estimation of the expectation values regardless of the form of the data distribution, and simplifies the algorithm considerably. It is allowed because the rank transformation is invertible and the maximum correlation is, by definition, invariant under invertible transformations [34].

IV. Prediction for the Sum of the C_1-C_4 Contents of the 35-70°C Fraction

The sum of the C_1-C_4 contents of the 35-70°C fraction is obtained and collected from a real industrial refinery for a period of about four months from July 07, 2022 to October 10, 2022 with the sampling frequency of three laboratory probes every day. The data was processed, i.e. outliers were removed, before the soft sensor was built. Thus, 307 labeled samples are collected. The first data segment from 1 to 246 is selected as the training data, and the remaining points from 247 to 307 are used for the test data set. In other words, the training and testing sample sizes for this case study are 246 and 61 data points, respectively.

It is important to note that the training data segment must contain a wide range of data (over the full range of plant operation) to obtain a robust soft sensor.

Five process variables, which are listed in Table 1, were used to predict the y values. We applied the ACE algorithm to training sample and the results for individual transformations of $\hat{\theta}(y)$, and $\{\hat{\phi}_i(x_i), i = 1, \cdots 5\}$ are plotted in Fig. 2.

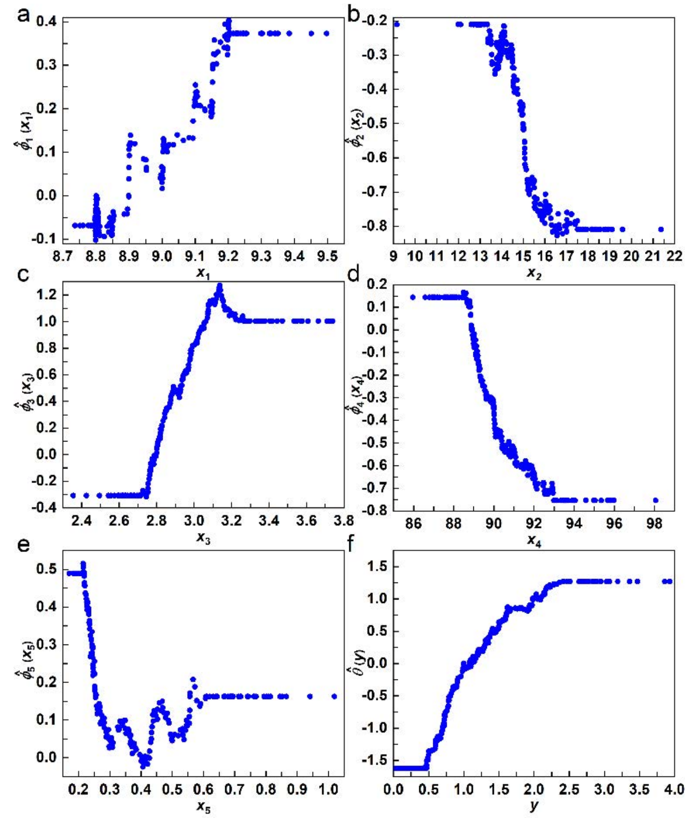

Fig. 2. Individual ACE optimal transformations ($\hat{\phi}_i(x_i)$) of training dataset as a function of the input variables: (a) the top pressure of the C-1; (b) the distillate flow from C-1; (c) the top pressure of the C-3; (d) the bottom temperature of the C-3; (e) the ratio of the reflux flow to C-3 to the feed flow to C-3. (f) Optimal transformation ($\hat{\theta}(y)$), which is a sum of $\hat{\phi}_i(x_i)$.

Fig. 2 illustrates the optimal transformations determined by ACE for the C-1 column top pressure and distillate flow, the C-3 column top pressure, the C-3 column bottom temperature, the ratio of reflux flow to the C-3 column to the C-3 column feed flow, and the total C1-C4 content in the C-3 column bottom stream. Although the ACE algorithm provides nonparametric optimization of the dependent and independent variables, it does not result a computational model for these variables. However, the optimal data transformations can be fitted, for example by simple polynomials to predict the dependent variable [35].

After applying the ACE algorithm, the dependencies of the transformed output on each of the transformed inputs are evaluated, which could be described as linear or nonlinear functions. Nonlinear functions $\hat{\theta}(y)$ and $\{\hat{\phi}_i(x_i), i = 1, \cdots 5\}$ are approximated to obtain their analytical form and fit results of the operating modes of the industrial unit.

Using the plots of Fig. 3 we obtained the analytic functions $\hat{\theta}(y)$ and $\{\hat{\phi}_i(x_i), i = 1, \cdots 5\}$:

$$\hat{\phi}_1(x_1) = 0.4 - \frac{0.5}{\left(1 + \frac{x_1}{9}\right)^{94}}$$

$$\hat{\phi}_2(x_2) = -0.8 + \frac{0.6}{\left(1 + \frac{x_2}{15}\right)^{38}}$$

$$\hat{\phi}_3(x_3) = 1 - \frac{1.3}{\left(1 + \frac{x_3}{3}\right)^{47}}$$

$$\hat{\phi}_4(x_4) = -0.7 + \frac{0.9}{\left(1 + \frac{x_4}{90}\right)^{124}}$$

$$\hat{\phi}_5(x_5) = 5 - 41x_5 + 131x_5^2 - 196x_5^3 + 140x_5^4 - 38x_5^5$$

and

$$\hat{\theta}(y) = 1.4 - \frac{3.1}{(1+y)^3}.$$

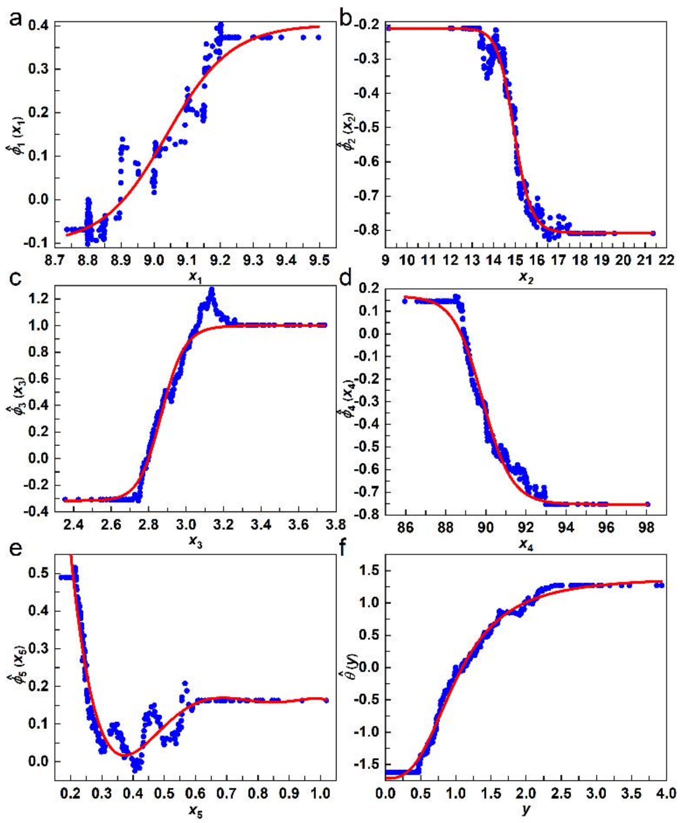

Fig. 3. Estimation of the optimal transformations for $\hat{\phi}_i(x_i)$ depending on the input variables: (a) the top pressure of the C-1; (b) the distillate flow from C-1; (c) the top pressure of the C-3; (d) the bottom temperature of the C-3; (e) the ratio of the reflux flow to C-3 to the feed flow to C-3. (f) Estimation of the optimal transformations $\hat{\theta}(y)$.

Fig. 4 shows the estimated sum of C_1-C_4 content in the bottom stream of the C-3 column using multiple linear regression – MLR (based on the method of least squares) as well as using the ACE approach.

To compare the accuracy of different methods for soft sensors design, the coefficient of determination (R^2) and the root mean-square error (*RMSE*) were chosen as criteria for estimation of the predicting capability performance of the soft sensors. These criteria are defined as follows:

$$R^2 = 1 - {\sum_n^N (y_n - \hat{y}_n)^2} \Big/ {\sum_n^N (y_n - \bar{y}_n)^2}, \quad (8)$$

$$RMSE = \sqrt{\sum_n^N (y_i - \hat{y}_i)^2 \Big/ N} \quad (9)$$

where N is the number of training or test samples, and y_n and $\hat{y}_n$, $n = \overline{1, N}$ are the actual and predicted values, respectively.

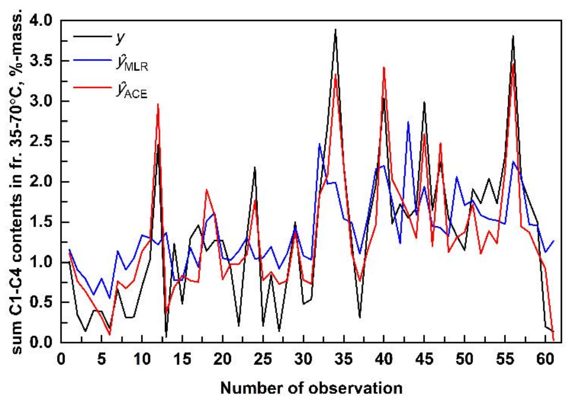

Fig. 4. Prediction result of the two different soft sensors.

Results obtained for a test sample with the size of 61 observations are presented in Table 2.

TABLE II. MODEL PERFORMANCE FOR THE INDUSTRIAL CASE

Method	Test sample	
	R^2	*RMSE*
MLR	0.4299	0.6736
ACE	0.7930	0.4059

The *RMSE* for multiple linear regression is 0.67 and 0.41 for nonlinear regression. The sum of C_1-C_4 estimated with nonlinear regression is very close to the original, the *RMSE* is up to 39 times lower than those obtained with MLR.

The nonlinear regression allows a particularly good prediction of the values in the range of the strong increase of the concentration of the sum of the C_1-C_4 hydrocarbons. Of great importance is that the ACE algorithm used for nonparametric soft sensor development has resulted in an improvement in estimation capability, namely an increase in R^2 and a decrease in *RMSE* by 100×(0.79-0.43)⁄0.43≈84% and 100×(0.67-0.41)⁄0.67≈39%, respectively.

V. CONCLUSION

The main objective of this paper was to test whether it is possible to predict the sum of C_1-C_4 hydrocarbons in the 35°C fraction using a nonlinear regression analysis. The data analyzed were obtained from the real industrial unit. The nonlinear input and response functions were first estimated on a training data set using the ACE algorithm and then tested on an independent test data set. The implementation of the ACE algorithm resulted in even higher soft sensor accuracy, which was independently confirmed for the industrial case compared to multiple linear regression.

REFERENCES

[1] N. Sifakis, N. Sarantinoudis, G. Tsinarakis, C. Politis, and G. Arampatzis, “Soft Sensing of LPG Processes Using Deep Learning,” J. Sens., vol. 23, no. 18, p. 7858, 2023.

[2] S. N. Sembodo, N. Effendy, K. Dwiantoro, and N. Muddin, “Radial basis network estimator of oxygen content in the flue gas of debutanizer reboiler,” Int. J. Electr. Comput. Eng., vol. 12, no. 3, pp. 3044–3050, 2022.

[3] X. Yuan, L. Li, Y. A. Shardt, Y. Wang, and C. Yang, “Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development,” IEEE Trans. Ind. Electron., vol. 68, no. 5, pp. 4404–4414, 2020.

[4] Y. Jiang, S. Yin, J. Dong, and O. Kaynak, “A review on soft sensors for monitoring, control, and optimization of industrial processes,” IEEE Sens. J., vol. 21, no. 11, pp. 12868–12881, 2020.

[5] S. A. Samotylova, A. Y. Torgashov, and A. A. Goncharov, “Structural identification of soft-sensor of MTBE unit,” 2017 Int. Conf. Ind. Eng. Appl. Manuf., pp. 1–5, May 2017.

[6] L. Fortuna, S. Graziani, A. Rizzo, M. G. Xibilia, Soft Sensors for Monitoring and Control of Industrial Processes. Switzerland: Springer, 2007, p. 284.

[7] J. A. Romagnoli and M. C. Sanchez, Data Processing and Reconciliation for Chemical Process Operations. CA:Academic Press, 1999, p. 270.

[8] S. Khatibisepehr and B. Huang, “Dealing with irregular data in soft sensors: bayesian method and comparative study,” Ind. Eng. Chem. Res., vol. 47, pp. 8713–8723, 2008.

[9] M. K. Asy’ari, R. D. Noriyati, and K. Indriawati, “Soft sensor design of solar irradiance using multiple linear regression,” 2019 Int. Sem. Intel. Technol.Appl., pp. 56–60, August 2019.

[10] B. Berger and F. Rauscher, “Robust Gaussian process modelling for engine calibration,” Proc. 7th Vienna Int. Conf. Math. Modelling, pp. 159–164, 2012.

[11] S. Cateni, V. Colla, and M. Vannucci, “Outlier detection methods for industrial applications,” in Advances in Robotics Automation and Control, Rijeka, Croatia: InTech, 2008, pp. 265–282.

[12] H. Feng and H. Zhang, “Soft sensor technique based on robust SVM,” 2005 Int. Conf. Neural Networks and Brain, vol. 3, pp. nil18–1707, October 2005.

[13] G. Khalaf and G. Shukur, “Choosing ridge parameter for regression problems,” Commun. Statist. Theory Methods, vol. 34, no. 5, pp. 1177–1182, May 2005.

[14] C. Saunders, A. Gammerman, and V. Vovk, “Ridge regression learning algorithm in dual variables,” Proc. 15th Int. Conf. Machine Learning, pp. 515–521, 1998.

[15] X. Yan, “Modified nonlinear generalized ridge regression and its application to develop naphtha cut point soft sensor,” Comput. Chem. Eng., vol. 32, pp. 608–621, 2008.

[16] Y. Wang and R. Chen, “A soft-sensor model for NOx concentration based on Enet-GPR,” 10th Data Driven Control Learning Systems Conf., pp. 236–241, May 2021.

[17] F. A. A. Souza and R. Araújo, “Mixture of partial least squares experts and application in prediction settings with multiple operating modes,” Chemom. Intell. Lab. Syst., vol. 130, pp. 192–202, 2014.

[18] X. Yuan, Z. Ge, Z. Song, Y. Wang, C. Yang, and H. Zhang, “Soft sensor modeling of nonlinear industrial processes based on weighted probabilistic projection regression,” IEEE Trans. Instrum. Meas., vol. 66, no. 4, pp. 837–845, 2017.

[19] X. Yuan, Z. Chen, and Y. Wang, “Probabilistic nonlinear soft sensor modeling based on generative topographic mapping regression,” IEEE Access, vol. 6, pp. 10445–10452, 2018.

[20] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald, “Data-based identifiability analysis of non-linear dynamical models,” Bioinf., vol. 23, no. 19, pp. 2612–2618, 2007.

[21] A. Di Ciaccio, G. E. Montanari, and others, “Non-parametric methods for data-mining applications,” Atti della XLI Riunione Scientifica della Societa Italiana di Statistica. Sessioni Plenarie e Sessioni Specializzate. CLEUP-Padova, pp. 339–348, 2002.

[22] S. Birle, M. A. Hussein, and T. Becker, “Fuzzy logic control and soft sensing applications in food and beverage processes,” Food Control, vol. 29, no. 1, pp. 254–269, 2013.

[23] E. Y. Nagai and L. V. R. de Arruda, “Soft sensor based on fuzzy model identification,” IFAC Proc. Vol., vol. 38, no. 1, pp. 127–132, 2005.

[24] X. Huo, K. Hao, L. Chen, X.-S. Tang, T. Wang, and X. Cai, “A dynamic soft sensor of industrial fuzzy time series with propositional linear temporal logic,” Expert Syst. Appl., vol. 201, p. 117176, 2022.

[25] I. Mendia, S. Gil-López, I. Landa-Torres, L. Orbe, and E. Maqueda, “Machine learning based adaptive soft sensor for flash point inference in a refinery realtime process,” Results Eng., vol. 13, p. 100362, 2022.

[26] M. Mowbray, H. Kay, S. Kay, P. Castro Caetano, A. Hicks, C. Mendoza, A. Lane, P. Martin, and D. Zhang, “Probabilistic machine learning based soft-sensors for product quality prediction in batch processes,” Chemom. Intell. Lab. Syst., vol. 228, p. 104616, 2022.

[27] P. Kadlec, R. Grbić, and B. Gabrys, “Review of adaptation mechanisms for data-driven soft sensors,” Comput. Chem. Eng., vol. 35, no. 1, pp. 1–24, 2011.

[28] O. Y. Snegirev and A. Y. Torgashov, “Adaptation of the structure and parameters of nonlinear soft sensors by the example of an industrial reactive distillation process,” Autom. Remote Control, vol. 82, pp. 1774–1786, 2021.

[29] P. Angelov, D. Filev, “An approach to on-line identification of evolving Takagi-Sugeno models”, IEEE Trans. on Systems, Man and Cybernetics, vol. 34, no 1, pp. 484–498, 2004.

[30] G. E. P. Box and P. W. Tidwell, “Transformation of the independent variables,” Technometrics, vol. 4, no. 4, pp. 531–550, 1962.

[31] G. E. P. Box and W. J. Hill, “Correcting inhomogeneity of variance with power transformation weighting,” Technometrics, vol. 16, no. 3, pp. 385–389, 1974.

[32] L. Breiman and J. H. Friedman, “Estimating optimal transformations for multiple regression and correlation,” J. Am. Stat. Assoc., vol. 80, no. 391, pp. 580–598, 1985.

[33] D. Wang and M. Murphy, “Estimating optimal transformations for multiple regression using the ACE algorithm,” J. Data Sci., vol. 2, no. 4, pp. 329–346, 2004.

[34] N. Wessel, J. Konvicka, F. Weidermann, S. Nestmann, R. Neugebauer, U. Schwarz, A. Wessel, J. Kurths, “Predicting thermal displacements in modular tool systems,” Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 14, no. 1, pp. 23–29, 2004.

[35] A. Malallah, R. Gharbi, and M. Algharaib, “Accurate estimation of the world crude oil PVT properties using graphical alternating conditional expectation,” Energy & Fuels, vol. 20, no. 2, pp. 688–698, 2006.