

Klimchenko, Vladimir; Torgashov, Andrei; Shardt, Yuri A.W.; Yang, Fan

Multi-output soft sensor with a multivariate filter that predicts errors applied to an industrial reactive distillation process




Original published in: Mathematics. - Basel : MDPI. - 9 (2021), 16, art. 1947, 14 pp.
Original published: 2021-08-15
ISSN: 2227-7390
DOI: [10.3390/math9161947](https://doi.org/10.3390/math9161947)
[Visited: 2022-03-04]



This work is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Article

Multi-Output Soft Sensor with a Multivariate Filter That Predicts Errors Applied to an Industrial Reactive Distillation Process

Vladimir Klimchenko ¹, Andrei Torgashov ^{1,*} , Yuri A. W. Shardt ²  and Fan Yang ³ 

¹ Process Control Laboratory, Institute of Automation and Control Process FEB RAS, 5 Radio Str., Vladivostok 690041, Russia; volk@iacp.dvo.ru

² Department of Automation Engineering, Technical University of Ilmenau, 99084 Ilmenau, Germany; yuri.shardt@tu-ilmenau.de

³ Beijing National Research Center for Information Science and Technology, Department of Automation, Tsinghua University, Beijing 100084, China; yangfan@tsinghua.edu.cn

* Correspondence: torgashov@iacp.dvo.ru; Tel.: +7-423-231-02-02

Abstract: The paper deals with the problem of developing a multi-output soft sensor for the industrial reactive distillation process of methyl tert-butyl ether production. Unlike the existing soft sensor approaches, this paper proposes using a soft sensor with filters to predict model errors, which are then taken into account as corrections in the final predictions of outputs. The decomposition of the problem of optimal estimation of time delays is proposed for each input of the soft sensor. Using the proposed approach to predict the concentrations of methyl sec-butyl ether, methanol, and the sum of dimers and trimers of isobutylene in the output product in a reactive distillation column was shown to improve the results by 32%, 67%, and 9.5%, respectively.

Keywords: soft sensing; multivariate filter; reactive distillation



Citation: Klimchenko, V.; Torgashov, A.; Shardt, Y.A.W.; Yang, F.

Multi-Output Soft Sensor with a Multivariate Filter That Predicts Errors Applied to an Industrial Reactive Distillation Process.

Mathematics **2021**, *9*, 1947. <https://doi.org/10.3390/math9161947>

Academic Editor: Michal Fečkan

Received: 20 June 2021

Accepted: 13 August 2021

Published: 15 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the size and complexity of industrial systems increases, there is a need to accurately measure most process variables. Unfortunately, not all variables can be accurately measured using online hard sensors. For certain variables, such as concentration or density, the only accurate measurements can be obtained by manually taking samples and analyzing them in a laboratory. One solution to this problem is the development of soft sensors, which take the easy-to-measure variables and create models to predict the hard-to-measure variables [1].

All soft sensor systems consist of a process model that takes the easy-to-measure variables and provides an estimate of the hard-to-measure variables. These models can be constructed using methods ranging from linear regression to principal component analysis and support vector machines. Although the main focus has been on the development of the soft sensor models [2–5], advanced soft sensor systems have also a bias update term that can take any slowly sampled information to update the soft sensor prediction [1]. This bias update term is normally designed as some function of the difference between the predicted and measured values [6]. Of note, it should be mentioned that the measured values are often sampled very slowly and with considerable time delay. This means that during the points at which there are no updates, the previously available bias value is used. When such a system is properly designed, it can provide good tracking of the process, i.e., the predicted and measured values are close to each other.

Recently, it has been suggested that instead of only using the available slowly sampled data for updating the bias term, it should be possible to also model the historical errors and use them to predict the future errors [7]. It has been shown that such an approach can improve the overall performance of the soft sensor system. However, there still remain issues with how best to model and implement this predictive bias update term.

Furthermore, there are issues with incorporating time delays into this approach since they will greatly increase the size of the required search space.

Therefore, this paper will examine the development of a predictive bias update term for a nonlinear system using dimension reduction. The proposed approach will be tested using data from an industrial reactive distillation column that produces methyl tert-butyl ether (MTBE).

2. Background

Consider the soft sensor system shown in Figure 1, where u_t is the input, y_t is the measured (true) output, $\hat{y}_{m,t}$ the predicted soft sensor value, $\hat{y}_{\alpha,t}$ and $\hat{y}_{\beta,t}$ are intermediate soft sensor values, G_p is the true process, \hat{G}_p is the soft sensor process model, and G_B is the bias update term. It can be noted that purpose of the bias update term is to take the information from the measured values and correct the output of the soft sensor system. This comes primarily from the unknown disturbances and the inherent plant-model mismatch.

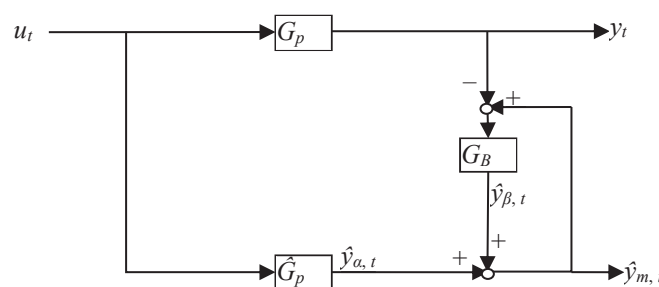


Figure 1. Soft sensor system of interest [1].

Another approach to this problem is to re-arrange the bias update term so that it contains a predictive model that can predict the errors between the measured and predicted values. This re-arrangement is shown in Figure 2, where the predicted value from the soft sensor is corrected based on the modeled errors of the system. The question becomes how to design this model so that the best predictions can be obtained.

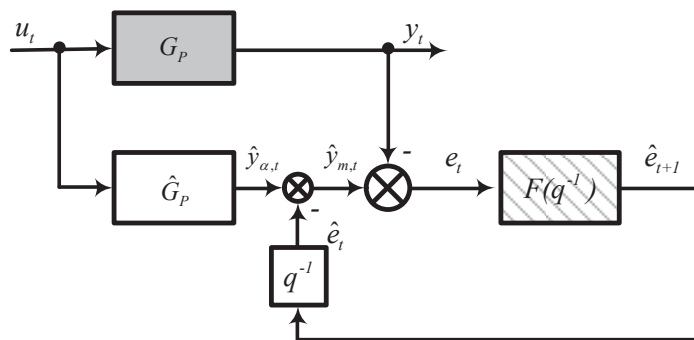


Figure 2. Bias update term as a predictive model with feedback: —plant, —predictive model.

For prediction of time series, the Box-Jenkins methodology is traditionally used, according to which the time series model is found in the class of autoregressive-moving average (ARMA) models, i.e., is considered a rational algebraic function of the backward shift operator. The flexibility of the ARMA class makes it possible to find parsimonious models, i.e., the adequacy of the evaluated model is achieved with a small number of estimated parameters. Since this property is especially important for empirical models, the Box-Jenkins methodology is widely used to solve various practical problems. This approach is adopted in this paper.

In industrial processes, where it is desired to implement the model on programmable logic control (PLC) units, the complexity of the model \hat{G}_p can be an issue. Therefore, this paper will consider a simple model for \hat{G}_p of the form

$$y_t = b_0 + bx_t + e_t \tag{1}$$

where b are the parameters to be estimated and x_t is the input(s). Model (1) can be improved by taking into account possible delays of the output variables relative to inputs. Consider the following model for a multi-output soft sensor

$$y_{t,m} = b_m u_m(t, \tau_m) + e_{t,m} \tag{2}$$

where $t = 1, 2, \dots, n$; $m = 1, 2, 3$ (the number of outputs m is given by the industrial production team and reflecting the key quality indices of MTBE product). Vector $b_m = (b_{m,1}, b_{m,2}, \dots, b_{m,10})$ is a row vector of unknown coefficients; $\tau_m = (\tau_{m,1}, \tau_{m,2}, \dots, \tau_{m,10})$ is a row vector of unknown time delays; $u_m(t, \tau_m) = (u_{t,m,1}, u_{t,m,2}, \dots, u_{t,m,10})^T$; $u_{t,m,k}$ is the measurement of the x_k value at time $t - \tau_{m,k}$ with $k = 1, 2, \dots, 10$. Please note that it has been assumed here that the maximal time delay is 10 samples and justified from the industrial process dynamics point of view. However, it can easily be extended to arbitrary values.

Solving model (2) by minimizing the mean squared error (MSE) gives an estimate for the unknown parameters \hat{b}_m and $\hat{\tau}_m$. The MSE depends not only on the coefficients b_m , but also on the delays τ_m , i.e.,

$$D_{em}(b_m, \tau_m) = \frac{1}{n} \sum_{t=1}^n \{y_{tm} - b_m u_m(t, \tau_m)\}^2, \quad m = 1, 2, 3 \tag{3}$$

Thus,

$$(\hat{b}_m, \hat{\tau}_m) = \arg \min_{b_m, \tau_m} D_{em}(b_m, \tau_m). \tag{4}$$

Please note that if $D_{em}(b_m^*, \tau_m^*) = \min_{b_m, \tau_m} D_{em}(b_m, \tau_m)$, then $D_{em}(b_m^*, \tau_m^*) = \min_{b_m} D_{em}(b_m, \tau_m^*)$.

Consequently,

$$\min_{b_m, \tau_m} D_{em}(b_m, \tau_m) = \min_{\tau_m} \left\{ \min_{b_m} D_{em}(b_m, \tau_m) \right\} = \min_{\tau_m} D_{em}(\hat{b}_m, \tau_m) \tag{5}$$

Furthermore, the estimates \hat{b}_m are found using standard regression analysis which gives

$$\hat{b}_m = \left\{ \left(\mathbf{U}_m^T \mathbf{U}_m \right)^{-1} \mathbf{U}_m^T \mathbf{Y}_m \right\}^T, \quad m = 1, 2, 3 \tag{6}$$

where \mathbf{Y}_m is the m -th column of the matrix \mathbf{Y} ; \mathbf{U}_m is a matrix with dimension $n \times 10$, whose t -th row is the row $u_m(t, \tau_m)^T$.

Since all variables are measured at discrete moments in time, the gradient descent methods cannot be directly applied to minimize the objective function $D_{em}(\hat{b}_m, \tau_m)$ for the argument τ_m . However, this difficulty can be avoided by calculating D_{em} for any values of the elements of the vector τ_m by interpolating between the nearby nodes of the discrete grid. Interpolation with a large search space dimension is a difficult problem. Among the various characteristics of the algorithms used, such properties as visibility and relative simplicity come to the fore. Therefore, in this situation, the most preferable is the polynomial interpolation.

2.1. Error Modeling

If the $e_{t,m}$ error were known at time $t - 1$, then using Equation (2), it would be possible to predict the $y_{t,m}$ variable with absolute accuracy. Unfortunately, the $e_{t,m}$ error is not known in advance, but it can be predicted using any statistical patterns found in the sequence $e_{1,m}, e_{2,m}, \dots$. This error prediction can be used as a correction to model (2) as shown in Figure 2, therefore improving the prediction accuracy of the $y_{t,m}$ output variable. To evaluate a predictive model for the sequence $e_{1,m}, e_{2,m}, \dots$, let us consider the class of ARMA models. Let us introduce the predicted process as the output of an invertible linear filter, called a shaping filter, driven by white noise, i.e., a process with a constant spectral density. In this case, the transfer function of the shaping filter is considered a rational algebraic function of the backward shift operator, i.e.,

$$e_t = \frac{\prod_{l=1}^{N_n} (1 - H_l q^{-1})}{\prod_{k=1}^{N_d} (1 - G_k q^{-1})} \varepsilon_t \tag{7}$$

where ε_t and e_t are values of the input and output processes of the shaping filter at time t ; N_n is the order of the moving average; N_d is the order of the autoregressive component; H_l, G_k are constants (generally speaking, complex-valued); and q^{-1} is the backshift operator. The stationarity and invertibility conditions, which are necessary to predict the e_t process, are [8]

$$|G_k| < 1, k = 1, \dots, N_d; |H_l| < 1, l = 1, \dots, N_n \tag{8}$$

The flexibility of the ARMA class provides the possibility of finding parsimonious models, i.e., the adequacy of the constructed model is achieved with a relatively small number of estimated parameters. Since this property is especially important for empirical models, the models with the structure given in Equation (7) and their variants are widely used for solving practical problems.

The filter for predicting the e_t process can be found using the prediction error method (PEM) [9]. Expanding the brackets in Equation (7) gives

$$e_t = \frac{(1 - \theta_1 q^{-1} - \dots - \theta_{N_n} q^{-N_n})}{(1 - \eta_1 q^{-1} - \dots - \eta_{N_d} q^{-N_d})} \varepsilon_t \tag{9}$$

where θ_l and η_k are the model parameters. It is assumed that the polynomials in the numerator and denominator have no common roots, since otherwise it would be possible to reduce the common multipliers in the numerator and denominator of Equation (7).

The PEM function finds the parameter values that minimize the predictive MSE of the e_t process for given polynomial orders (N_n, N_d) and the initial estimates of the parameters θ_l and η_k . It is possible to choose suitable orders of the polynomials based on sample estimations of the spectral density of the considered process. Recall that the frequency response of the shaping filter is the value of Equation (7) on a circle of unit radius centered on the origin and the spectral density $S(\omega)$ of the output process e_t is equal to the product of the variance of the input process and the square of the frequency response modulus, i.e., [10]

$$S(\omega) = \sigma_\varepsilon^2 \frac{\prod_{l=1}^{N_n} (1 - H_l e^{-j\omega}) \prod_{l=1}^{N_n} (1 - \bar{H}_l e^{j\omega})}{\prod_{k=1}^{N_d} (1 - G_k e^{-j\omega}) \prod_{k=1}^{N_d} (1 - \bar{G}_k e^{j\omega})}, \tag{10}$$

where σ_ε^2 is the variance of random process ε_t and H_l and G_k are the complex conjugates of the constants H_l and G_k . Furthermore, since we desire that our filter be invertible, it follows that for the model

$$\varepsilon_t = \frac{\prod_{k=1}^{N_d} (1 - G_k q^{-1})}{\prod_{l=1}^{N_n} (1 - H_l q^{-1})} e_t \tag{11}$$

the e_t process is invertible if the absolute values of all the H_l constants are less than one. Similarly, if the absolute values of all the G_k constants is less than one, then the e_t process is stationary [8]. Thus, although multiple processes can have the same spectral density, there is only one that is both stationary and invertible.

Once the general model has been obtained, we can rewrite it as an infinite impulse response model, i.e.,

$$e_t = \varepsilon_t + \sum_{k=1}^{\infty} \psi_k \varepsilon_{t-k} \tag{12}$$

where ψ is an impulse response coefficient. Since we know that the general model converges [8], it follows that we only need a finite number of terms in Equation (12). Furthermore, we note that

$$e_{t-i} = \varepsilon_{t-i} + \sum_{k=1}^{\infty} \psi_k \varepsilon_{t-i-k} \tag{13}$$

which implies that for any positive i the random variables ε_t and e_{t-i} are uncorrelated (since the process ε_t is white noise). Therefore, successively multiplying both sides of Equation (12) by the values of the corresponding process at delays i and taking expectations, we obtain equations for finding the initial estimates of the parameters that involve the covariances of the errors for different lags [10]. Obviously, since the true covariances are not known, they will need to be replaced by the sample estimates. This method of estimating the coefficients does not lead to too large error as long as the absolute values of the parameters of model (7) are not too close to the boundary of unit circle centered on the origin. Thus, it is possible to design the required filter.

2.2. Filter Design

Let $e_t = (e_{t,1}, e_{t,2}, \dots, e_{t,N})^T$ be an N -dimensional stationary process of the soft sensor's errors whose shaping filter transfer matrix is $F_0(q^{-1})$, i.e.,

$$e_t = F_0(q^{-1})\varepsilon_t \tag{14}$$

where q^{-1} is the backshift operator; $\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, \dots, \varepsilon_{t,N})^T$ is an N -dimensional vector of white noise; and $F_0(q^{-1}) = [f_{km}(q^{-1})]$ is an $N \times N$ matrix function, whose entries denoted as $f_{km}(q^{-1})$ are the rational transfer function from $\varepsilon_{t,m}$ to $e_{t,k}$. Thus, it is desired to construct the filter that will predict e_{t+1} given the past values.

Let $P(q^{-1})$ be the desired one-step ahead predictor transfer matrix, $\hat{e}_{t+1} = P(q^{-1})e_t$ the prediction of the vector e_{t+1} at time t , and $\tilde{\varepsilon}_{t+1} = e_{t+1} - \hat{e}_{t+1}$ the error of the prediction obtained with the aid of the filter $P(q^{-1})$. Then

$$\tilde{\varepsilon}_t = e_t - \hat{e}_t = e_t - q^{-1}\hat{e}_{t+1} = e_t - q^{-1}P(q^{-1})e_t = [I_N - q^{-1}P(q^{-1})]e_t \tag{15}$$

where I_N is identity matrix of order N . Consequently, the filter in the square brackets transforms the initial series into the prediction error series. If the random vector $\tilde{\varepsilon}_t$ includes components correlated with those of the vector $\tilde{\varepsilon}_{t-j}$ at some $j > 0$, we can predict the errors $\tilde{\varepsilon}_t$ using the known previous errors. Using those predictions as corrections to the $\tilde{\varepsilon}_t$ that were obtained, we could improve the accuracy of the predictions. Hence, in order to maximize the predictor accuracy, we must find a $P(q^{-1})$ such that the errors $\tilde{\varepsilon}_t$ are uncorrelated with the errors $\tilde{\varepsilon}_{t-j}$ at any $j > 0$ with some nonzero correlation between the

components of $\tilde{\varepsilon}_t$ (i.e., at $j = 0$) being admissible. In other words, the time series $\tilde{\varepsilon}_t$ must be N -dimensional white noise. Consequently, $I_N - q^{-1}P(q^{-1}) = F_0^{-1}(q^{-1})$, from which it follows that $P(q^{-1}) = q[I_N - F_0^{-1}(q^{-1})]$.

Thus, the predictor transfer matrix $P(q^{-1})$ can be expressed through the transfer matrix of the shaping filter $F_0(q^{-1})$. The matrix $F_0(q^{-1})$ can be found from

$$G(q^{-1}) = F_0(q^{-1})F_0^T(q), \tag{16}$$

where $G(q^{-1}) = [g_{km}(q^{-1})]$, $g_{km}(q^{-1})$ is the q -transform of the statistical estimate of the cross-covariance function of the time series $e_{t,k}$ and $e_{t,m}$ (in particular, when $m = k$, g_{mm} is a q -transform of the sample covariance function, i.e., the autocovariance generating function (AGF) of the time series e_{tm}).

The algorithm for finding $F_0(q^{-1})$ is simplified by decomposing it into N stages. At the k th stage, a shaping filter $F_k(q^{-1})$ of the k -dimensional process $(e_{t,1}, e_{t,2}, \dots, e_{t,k})^T$ is found. At this stage, the filter $F_{k-1}(q^{-1})$, found at the $(k-1)$ th stage, is used in order to transform the matrix $G_k(q^{-1}) = F_k(q^{-1})F_k^T(q)$ so that its transform contains nonzero elements in only one line, one column, and on the main diagonal. This technique substantially simplifies the procedure of spectral factorization (finding the matrix function $F_k(q^{-1})$) [11].

The proposed approach allows us to identify the vector time series transfer matrix without resorting to a complicated phase state representation. This advantage is used to obtain an adequate model with relatively few estimated parameters for the initial time series shaping filter $F_0(q^{-1})$. Simultaneously, the model for the transfer matrix of the inverse filter $F_0^{-1}(q^{-1})$, which transforms the initial time series into the white noise, is also found.

The algorithm for constructing both the shaping filter $F_0(q^{-1})$ and its inverse $F_0^{-1}(q^{-1})$ is described in [11]. Based on this algorithm, the sequence of prediction errors $\tilde{\varepsilon}_t$ should be N -dimensional white noise. However, since in practice, the true characteristics of the original process are not known, but only their estimates, containing inevitable statistical errors, in reality, the properties of the sequence $\tilde{\varepsilon}_t$ can be significantly different from the properties of white noise. Thus, to verify the optimality of the resulting model $P(q^{-1})$ of the predictive filter, a criterion is needed to test the hypothesis that the process $\tilde{\varepsilon}_t$ is N -dimensional white noise. To construct such a criterion, we can transform the process $\tilde{\varepsilon}_t$ in such a way that its spectral density matrix is diagonal. Such a transformation is achieved by means of a rotation of axes in the N -dimensional variable space $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_N$ [12]. Since the variances of these variables can be made equal to each other by normalization, without loss of generality, we suppose that spectral density matrix of the noise $\tilde{\varepsilon}_t$ is an $N \times N$ identity matrix I_N .

Consider a univariate sequence $\zeta_k = \tilde{\varepsilon}_{t-j,m}$, where $k = jN + m$. Please note that each pair couple (j, m) determines one k and each k determines one pair couple (j, m) . Consequently, $\tilde{\varepsilon}_t$ is multivariate white noise if and only if ζ_k is univariate white noise. It is known that the spectral density of univariate white noise is constant [8,13]. Thus, testing the hypothesis that $\tilde{\varepsilon}_t$ is multivariate white noise is reduced to testing the hypothesis on the constancy of the spectral density of a univariate sequence. This hypothesis can be tested using Kolmogorov's criterion [14].

Please note that only a time series containing prediction errors is used as the initial information for constructing a predictor with the proposed approach. Information about the model with which the predictions were obtained is not used. Therefore, this approach is applicable to any predictive model that involves errors, regardless of the specific properties of the model used.

2.3. Summary of the Proposed Approach

Thus, the proposed procedure for developing the model can be summarized as follows:

Step 1: Create an initial sample $u_t, y_t, t = 1, 2, \dots, K$. If the plant is already functioning then the initial sample consists of the historical values of u_t, y_t . Otherwise, the initial sample is forming during the trial period of the plant. The initial sample is divided into training and testing datasets.

Step 2: Based on the data included in the training sample, the coefficients and delays of the model given by Equation (2) are estimated via solving optimization problem (4).

Step 3: Based on the data included in the training sample, the errors for the model and the corresponding sample spectrum of errors are calculated.

Step 4: Based on the sample spectrum, the order of the ARMA model is selected in order to predict the unknown future error given the known current and past errors.

Step 5: The least squares method is used to find the values of the ARMA model parameters.

Step 6: The ARMA model obtained is used as the predictive filter $F(q^{-1})$ in the feedback loop of the compensator (bias update term) as shown in Figure 2.

Step 7: If the resulting soft sensor improves the accuracy of the prediction for the test sample then it can be recommended for practical use.

Please note that the obtained predictive filter model can be recommended for further use for the same plant on the data of which it was built. As for the approach, it will certainly be successful if the sequence of errors of the plant is a stationary (or close to it) process. In addition, the class of successful applicability of this approach can be extended to those plants, for whose errors it is possible to find an invertible transformation that brings the sequence of errors to a stationary process. The quality of the developed model should be checked on a test sample that was not used at the stage of the model training.

3. Industrial Application of the Proposed Method

Industrial methyl tert-butyl ether (MTBE) production occurs in a reactive distillation unit, as shown in Figure 3. The feed containing isobutylene and methanol (MeOH) enters the column. The distillate (D) is a lean butane-butylene fraction with a certain amount of MeOH. The raffinate is the heavy product MTBE that is withdrawn from the bottom part of the column. Table 1 shows the main process variables for the industrial unit. The goal is to develop a soft sensor for the prediction of the concentrations of methyl sec-butyl ether (MSBE), MeOH, and the sum of dimers and trimers of isobutylene (DIME) in the bottom product MTBE.

The measured values of output y_m and input x_k variables at the time moment t are denoted as $y_{tm}, x_{tk}; m = 1, 2, 3; k = 1, 2, \dots, 10; \text{ and } t = 1, 2, \dots, n$. The existing measurements may be used for development of a predictive model of the form

$$y_t = b_0 + bx_t + e_t, \quad t = 1, 2, \dots, n \tag{17}$$

where $y_t = (y_{t,1}, y_{t,2}, y_{t,3})^T; x_t = (x_{t,1}, x_{t,2}, \dots, x_{t,10})^T; b$ is a matrix of the model parameters $[b_{mk}]$ of dimension $3 \times 10; b_0 = (b_1, b_2, b_3)^T$ is a vector of the constant biases; $e_t = (e_{t,1}, e_{t,2}, e_{t,3})^T$ is a vector of the residuals, and the superscript T denotes the transpose. Since Equation (17) can be rewritten as

$$(y_t - \bar{y}) = b(x_t - \bar{x}) + e_t \tag{18}$$

where $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t, \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$, then expectations of all the elements of vectors y_t, x_t , and e_t , as well as biases vector b_0 , may be considered to be equal to zero without loss of generality.

Although the elements of matrix b are unknown, they are easily estimated using the ordinary least squares (OLS) method, which gives [10]

$$\hat{b} = \left\{ \left(X^T X \right)^{-1} X^T Y \right\}^T \tag{19}$$

where $X = [x_{tk}]; Y = [y_{tm}]; m = 1, 2, 3; k = 1, 2, \dots, 10; \text{ and } t = 1, 2, \dots, n$.

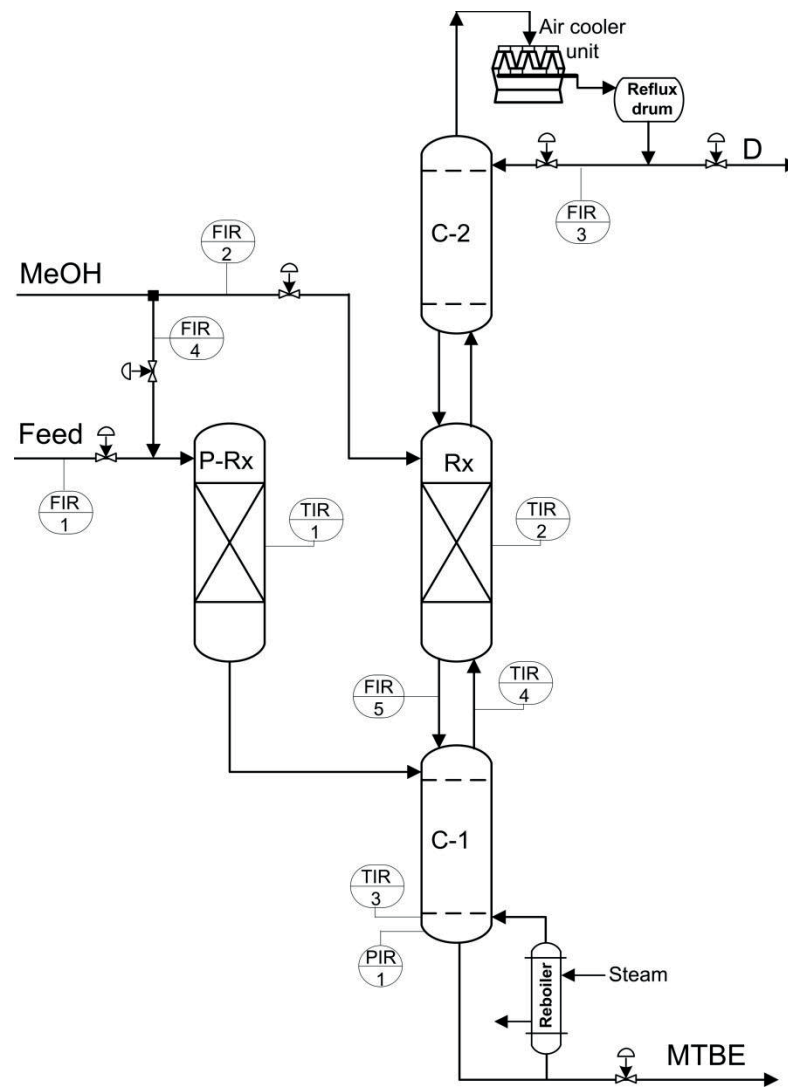


Figure 3. Reactive distillation unit of MTBE production.

Table 1. Soft sensor input and output variables.

Description of Process Variable	Notation	SS Variable
Feed flowrate, m ³ /s	FIR-1	x ₁
MeOH flowrate to Rx, m ³ /s	FIR-2	x ₂
Reflux flowrate, m ³ /s	FIR-3	x ₃
MeOH flowrate to P-Rx, m ³ /s	FIR-4	x ₄
Bottoms flowrate from Rx, m ³ /s	FIR-5	x ₅
Bottom pressure, MPa	PIR-1	x ₆
Temperature in P-Rx, K	TIR-1	x ₇
Temperature in Rx, K	TIR-2	x ₈
Bottom temperature, K	TIR-3	x ₉
Vapor flow temp. from C - 1, K	TIR-4	x ₁₀
Concentration of MSBE in MTBE, wt. %	-	y ₁
Concentration of MeOH in MTBE, wt. %	-	y ₂
Concentration of DIME in MTBE, wt. %	-	y ₃

For the training sample containing $n = 400$ measurements, the following estimates were obtained:

$$\bar{x} = (51.8154 \ 1.8747 \ 52.1154 \ 3.0859 \ 51.9866 \ 0.7580 \ 60.7100 \ 66.4516 \ 136.3077 \ 64.5725)^T$$

$$\bar{y} = (0.5440 \ 0.1461 \ 0.0595)^T$$

$$\hat{b} = \begin{pmatrix} -0.0151 & 0.2383 & -0.0342 & 0.1401 & 0.0476... \\ -0.0173 & 0.0794 & 0.0281 & 0.1191 & -0.0171... \\ -0.0080 & 0.1118 & -0.0061 & 0.0537 & 0.0134 \dots \\ -1.7361 & 0.0430 & -0.0012 & -0.1019 & 0.0388 \\ 2.9800 & -0.0093 & -0.0072 & -0.1333 & 0.0353 \\ 0.3490 & 0.0215 & -0.0011 & -0.0467 & 0.0098 \end{pmatrix}.$$

The estimated MSE vector for the model (17) is $(0.0094 \ 0.0095 \ 0.0021)^T$, while the vector of sample estimates of variances of the output variables is $(0.0321 \ 0.0184 \ 0.0047)^T$.

Let R_m^2 be a sample estimate of the coefficient of determination, i.e., the estimate of a fraction of variance of the dependent variable y_m explained by model (18), i.e.,

$$R_m^2 = 1 - \frac{D_{e,m}}{D_m} \tag{20}$$

where D_m is a sample estimate of the variance of the output variable y_m , $D_{e,m}$ is the mean squared value of the $e_{t,m}$ errors, and $m = 1, 2, 3$. This gives $R_1^2 = 0.7061$, $R_2^2 = 0.4822$, and $R_3^2 = 0.5467$.

Assuming a sampling time of one hour, the estimates of the delay vector $\hat{\tau}_1$ for predicting the output variable y_1 is

$$\hat{\tau}_1 = (4.83 \ 0 \ 2.00 \ 5.00 \ 1.83 \ 0 \ 2.00 \ 0.83 \ 1.00 \ 2.00)$$

and the estimate of the coefficient vector is equal to

$$\hat{b}_1 = (0.0002 \ 0.1341 \ -0.0360 \ 0.0064 \ 0.0451 \ -2.3289 \ 0.0519 \ -0.0029 \ -0.0819 \ 0.0442)$$

with $D_{e,1}(\hat{b}_1, \hat{\tau}_1) = 0.0091$.

Similarly, for variables y_2 and y_3 , we obtain

$$\hat{\tau}_2 = (0.33 \ 0.33 \ 1.67 \ 4.50 \ 0.50 \ 0.67 \ 0.33 \ 0.50 \ 0.50 \ 1.67)$$

$$\hat{b}_2 = (-0.0263 \ 0.1481 \ 0.0315 \ 0.1947 \ -0.0168 \ 3.4223 \ -0.0064 \ -0.0092 \ -0.1513 \ 0.0385); D_2(\hat{b}_2, \hat{\tau}_2) = 0.0088$$

$$\hat{\tau}_3 = (4.17 \ 0 \ 0.83 \ 4.33 \ 0.83 \ 0.50 \ 2.00 \ 0.67 \ 0.83 \ 1.00)$$

$$\hat{b}_3 = (-0.0021 \ 0.0811 \ -0.0070 \ -0.0016 \ 0.0130 \ 0.3795 \ 0.0259 \ -0.0015 \ -0.0455 \ 0.0098); D_3(\hat{b}_3, \hat{\tau}_3) = 0.0020$$

The sample estimate of the coefficient of determination to predict the output variable y_m denoted by R_{Lm}^2 is $R_{L1}^2 = 0.7160$; $R_{L2}^2 = 0.5200$; $R_{L3}^2 = 0.5726$.

The effect of delay accounting was evaluated on a test sample containing 167 measurements. As a result, the MSE of the predictions of output variables y_1 , y_2 and y_3 decreased by 23%, 10%, and 3%, respectively.

Now, let us consider modeling the error term. From the spectral density of the errors for $e_{t,1}$ and $e_{t,3}$ shown in Figures 4 and 5, it can be seen that the maximum within the interval $[0, 0.5]$ Hz indicates the presence in the denominator of the spectral density function $S(\omega)$ a factor $(1 - Ge^{-j\omega})$ with a complex-valued constant G . Since the sampling time is equal to 12 h, the frequency unit $1/(12 \text{ h})$ is used instead of Hz. However, for the practical application of the filter given by Equation (9), it is necessary that all the coefficients be real [8]. Therefore, the denominator of density $S(\omega)$ must contain a factor

$(1 - \bar{G}e^{-j\omega})$ along with a factor $(1 - Ge^{-j\omega})$. If the frequency response models for $e_{t, 1}$ and $e_{t, 3}$ processes are limited to these two factors (assuming the numerator is equal to one), then the corresponding spectral density of the second-order autoregressive process approximates well the sample estimates of the spectrum of $e_{t, 1}$ and $e_{t, 3}$ processes at different values of G . However, the insufficiently rapid decrease of the spectral density in the high-frequency region justifies the inclusion in the denominator of the model another multiplier with a real value of the constant G .

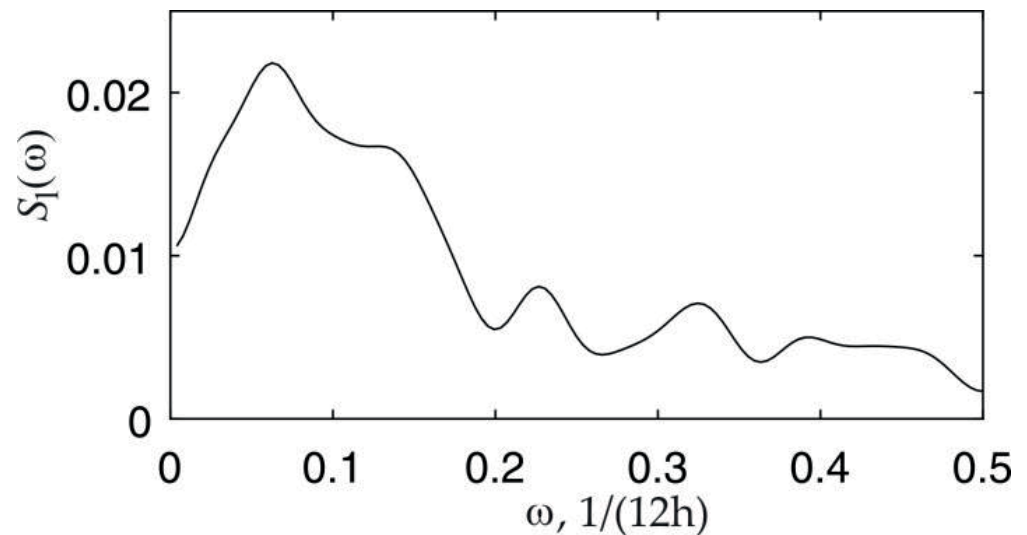


Figure 4. Sample spectrum of the process $e_{t, 1}$.

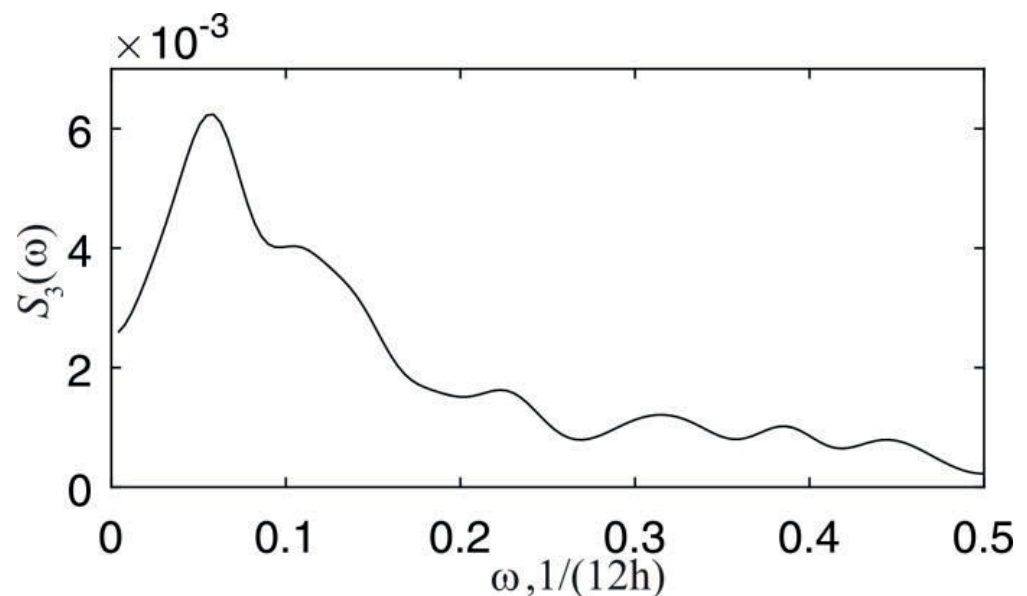


Figure 5. Sample spectrum of the process $e_{t, 3}$.

In Figure 6, which shows the spectral density for the $e_{t, 2}$ errors, the sample spectrum of this time series resembles the spectrum of a first-order autoregressive process [15–17]. However, we note that the stochastic process is not uniquely determined by its spectral density [8]. Therefore, as previously mentioned, we need to include two additional constraints that the resulting model be invertible and realizable. This will ensure that we have a unique model.

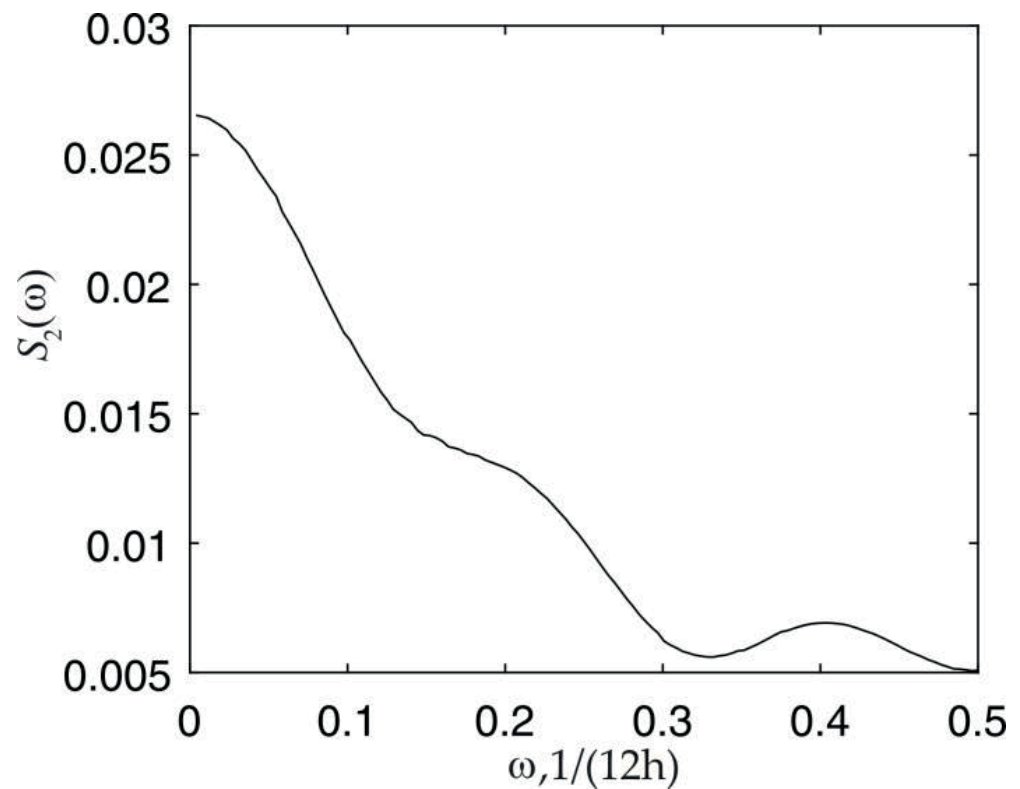


Figure 6. Sample spectrum of the process $e_{t, 2}$.

Based on the theoretical properties of the process, the error models are

$$\begin{aligned} e_{t, 1} - \eta_{11}e_{(t-1), 1} - \eta_{21}e_{(t-2), 1} - \eta_{31}e_{(t-3), 1} &= \varepsilon_{t, 1} \\ e_{t, 2} - \eta_{12}e_{(t-1), 2} &= \varepsilon_{t, 2} \end{aligned} \tag{21}$$

$$e_{t, 3} - \eta_{13}e_{(t-1), 3} - \eta_{23}e_{(t-2), 3} - \eta_{33}e_{(t-3), 3} = \varepsilon_{t, 3}$$

where η are the parameters to be determined. These parameters can be found using the approach presented in Section 2.2 by multiplying the finite impulse response model by the delayed errors and taking the expectations. For example, for e_1 , this gives

$$\gamma_i = \eta_{11}\gamma_{i-1} + \eta_{21}\gamma_{i-2} + \eta_{31}\gamma_{i-3}, \quad i = 1, 2, 3 \tag{22}$$

where $\gamma_i = \text{cov}(e_{t1}, e_{(t-i)1}) = \gamma_{-i}$.

For the process $e_{t, 1}$, the estimates of the coefficients η_{11} , η_{21} and η_{31} are, respectively, equal to 0.4131, -0.0093 , and -0.0528 . These values were used as the initial guesses passed to the PEM function. As a result of calculations, the model parameters were found to be: $\eta_{11} = 0.4175$, $\eta_{21} = 0.03234$, $\eta_{31} = -0.07026$. The initial value of coefficient η_{12} is 0.3748 and its final value is $\eta_{12} = 0.3758$.

Similarly, using Equation (22), the initial guesses were $\eta_{13} = 0.5142$, $\eta_{23} = -0.0507$, and $\eta_{33} = -0.0207$ to give final values of $\eta_{13} = 0.5151$, $\eta_{23} = -0.02676$, and $\eta_{33} = -0.03246$.

The performance of predictive filter models obtained from the analysis of the training dataset is validated using the testing sample. Figures 7–9 compare the predictions against the true values, where the solid line shows the true $e_{t, m}$ errors and the dashed line their predicted values for $m = 1, 2$, and 3. At the time point t on the x -axis, the corresponding error $e_{t, m}$ and the predicted error $\hat{e}_{t, m}$ computed at $t - 1$.

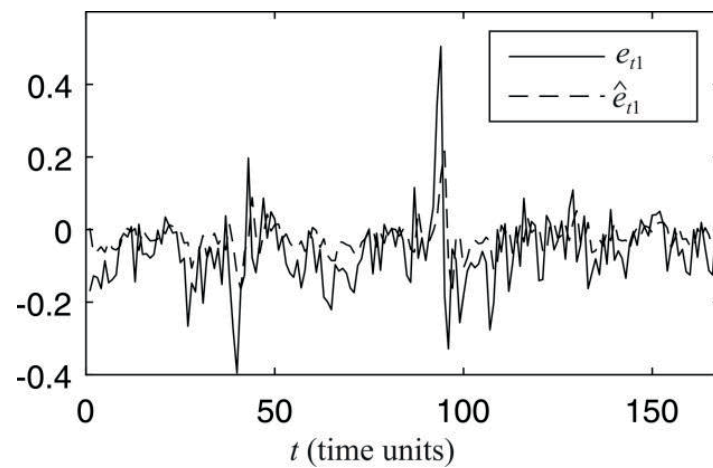


Figure 7. Prediction of the process $e_{t,1}$.

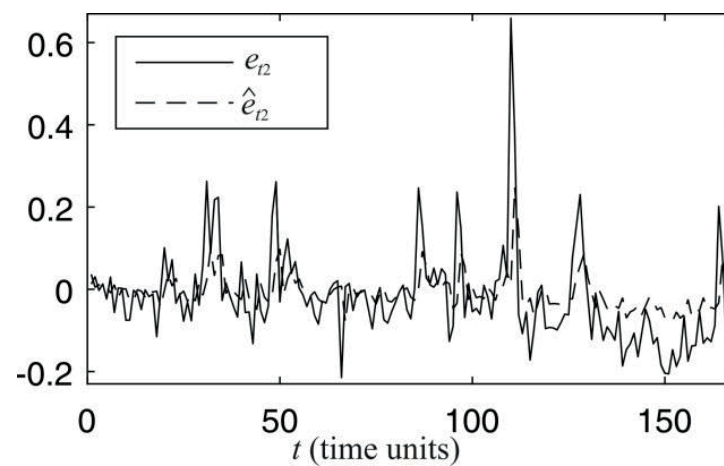


Figure 8. Prediction of the process $e_{t,2}$.

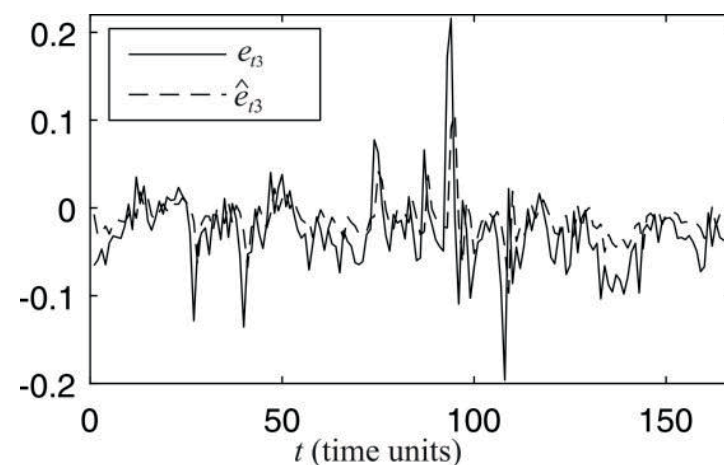


Figure 9. Prediction of the process $e_{t,3}$.

Figures 10–12 compare the performance of the soft sensors with the proposed filter for error prediction and a traditional method, in which adaptive bias term is calculated based on the moving window (MW) approach [18]. It can be seen that the filter provides better tracking of the process values, therefore improving the accuracy of the overall soft sensor system reducing the MSE of the output variables y_1 , y_2 , and y_3 by 32%, 67%, and 9.5%, respectively.

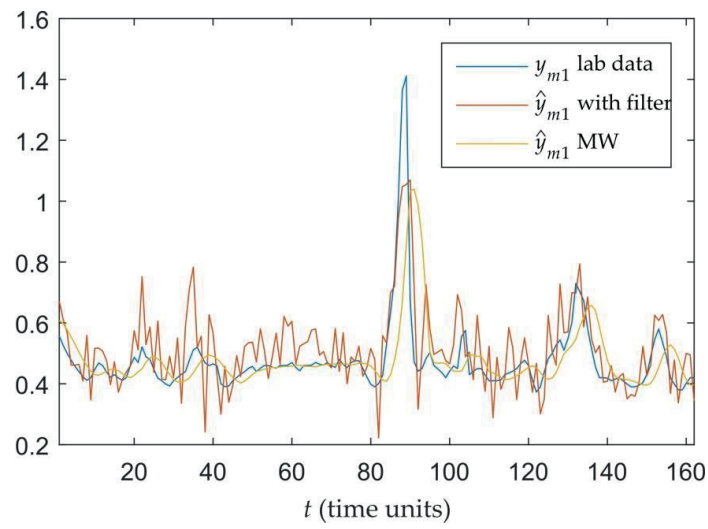


Figure 10. Estimation of y_{m1} .

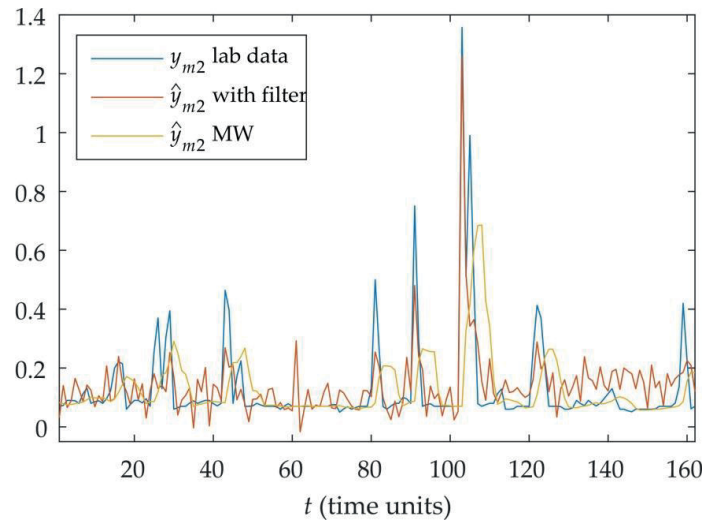


Figure 11. Estimation of y_{m2} .

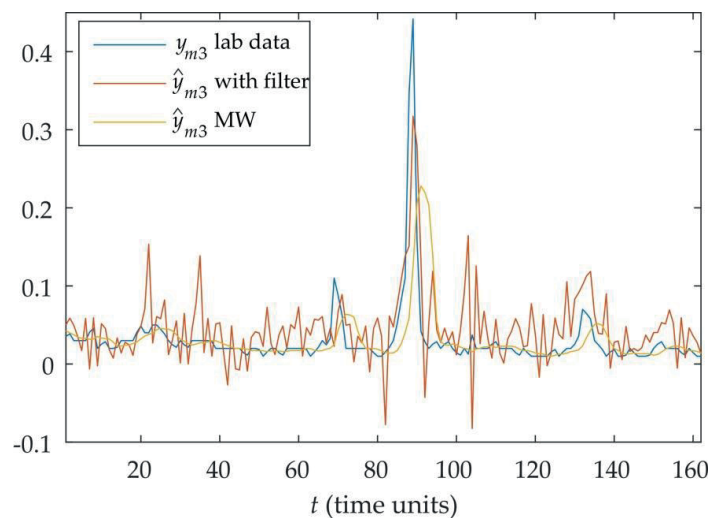


Figure 12. Estimation of y_{m3} .

4. Conclusions

This paper proposed a new approach to handling the bias update term in a soft sensor system. Rather than purely using available samples, the new bias update term seeks to predict what the errors will be in the future. Tests of this approach on a reactive distillation column show that the approach can handle the errors well. However, the predictive filters used only work for areas without serious disturbances or outliers.

Therefore, it makes sense to consider more complex models for the predictive filters including models with an additional component in the form of some flow, for example, Poissonian flow, of events (outliers). If the flow of outliers is added to the process model then the intensity of this flow needs to be estimated. In this case, the number of outliers in the training dataset should be sufficient to estimate the intensity of the flow of outliers with acceptable accuracy.

Author Contributions: Conceptualization, all; methodology, Y.A.W.S., A.T., V.K.; software, A.T.; validation, A.T., Y.A.W.S.; formal analysis, all; resources, F.Y., V.K.; writing—original draft preparation, Y.A.W.S., A.T.; writing—review and editing, all; funding acquisition, F.Y., V.K., Y.A.W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by RFBR and NSFC (grant numbers 21-57-53005 and 62111530057) and National Science and Technology Innovation 2030 Major Project (grant No.2018AAA0101604) of the Ministry of Science and Technology of China.

Data Availability Statement: Data can be obtained by contacting the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shardt, Y.A.W. *Data Quality Assessment for Closed-Loop System Identification and Forecasting with Application to Soft Sensors*; University of Alberta Press: Ed-monton, AB, Canada, 2012; Available online: <https://era.library.ualberta.ca/items/8382f12a-8960-4508-9ede-0679e021394b> (accessed on 2 January 2021).
2. Bakirov, R.; Babrys, B.; Fay, D. Multiple adaptive mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2017**, *96*, 42–54. [[CrossRef](#)]
3. Funatsu, K. Process control and soft sensors. In *T. Engel, & J. Gasteiger, Applied Chemoinformatics: Achievements and Future Opportunities*; Wiley-VCH: Weinheim, Germany, 2018; pp. 571–584.
4. Kim, S.; Kano, M.; Hasebe, S.; Takimi, A.; Seki, T. Long-term industrial applications of inferential control based on just-in-time soft-sensors: Economical impact and challenges. *Ind. Eng. Chem. Res.* **2013**, *52*, 12346–12356. [[CrossRef](#)]
5. Torgashov, A.; Skogestad, S. The use of first principles model for evaluation of adaptive soft sensor for multicomponent distillation unit. *Chem. Eng. Res. Des.* **2019**, *151*, 70–78. [[CrossRef](#)]
6. Griesing-Scheiwe, F.; Shardt, Y.A.; Pérez-Zuñiga, G.; Yang, X. Soft Sensor Design for Restricted Variable Sampling Time. *J. Process. Control.* **2020**, *92*, 310–318. [[CrossRef](#)]
7. Klimchenko, V.V.; Samotylova, S.A.; Torgashov, A.Y. Feedback in a predictive model of a reactive distillation process. *J. Comput. Syst. Sci. Int.* **2019**, *58*, 637–647. [[CrossRef](#)]
8. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; Wiley: Hoboken, NJ, USA, 2016.
9. Ljung, L. *System Identification*; Prentice Hall: Englewood Cliffs, NJ, USA, 1987.
10. Shardt, Y.A.W. *Statistics for Chemical and Process Engineers*; Springer: Berlin/Heidelberg, Germany, 2015.
11. Klimchenko, V.V. Decomposition of the multi-dimensional time series identification problem. *Autom. Remote. Control.* **2008**, *69*, 845–857. [[CrossRef](#)]
12. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*, 3rd ed.; John Wiley: New York, NY, USA, 2003.
13. Brillinger, D.R. *Time Series: Data Analysis and Theory*; SIAM: Philadelphia, PA, USA, 2001.
14. Marsaglia, G.; Tsang, W.W.; Wang, J. Evaluating Kolmogorov's Distribution. *Stat. Softw.* **2003**, *8*, 1–4. [[CrossRef](#)]
15. Hoff, J.C. *A Practical Guide to Box-Jenkins Forecasting*; Lifetime Learning Publications: Belmont, CA, USA, 1983.
16. Hannan, E.J.; Deistler, M. *Statistical Theory of Linear Systems*; John Wiley and Sons: New York, NY, USA, 1988.
17. Marple, S.L. *Digital Spectral Analysis*, 2nd ed.; Courier Dover Publications: Chicago, IL, USA, 2019.
18. Kadlec, P.; Grbic, R.; Gabrys, B. Review of adaptation mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2011**, *35*, 1–24. [[CrossRef](#)]