International Journal Of
Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

# Tail Gas quality Warning System in a Sulfur Recovery unit based on $H_2S$ and $SO_2$ Concentration Soft Sensor Utilizing Multi-State-Dependent Modeling Method

Fereshte Tavakoli Dastjerd[1] | Farhad Shahraki[*1] | Jafar Sadeghi[1] | Mir Mohammad Khalilipour[1]| Bahareh Bidar[1]

Center for Process Integration and Control (CPIC), Department of Chemical Engineering, University of Sistan and Baluchestan, Zahedan 98164, Iran.[1]

Corresponding author's email: fshahraki@eng.usb.ac.ir

**Article Info**

**Article type:**
Research Article

**Article history:**
Received: 13-August-2023
Received in revised form: 19-November-2023
Accepted: 23-November-2023
Published online: 10- December -2023

**Keywords**:
Pollutants,
Data-Driven Soft Sensor,
Sulfur Recovery Unit,
Multi-State-Dependent Parameter,
Generalized Random Walk

**ABSTRACT**

The design and development of data-driven soft sensors is important to predict the concentration of perilous pollutants in industry effluents to protect environmental health. The aim of this research is to design a tail gas quality warning system in the sulfur recovery unit (SRU) based on $H_2S$ and $SO_2$ concentration soft sensor utilizing multi-state-dependent modeling method. The SRU in the petrochemical plant of ERG PETROLI, located in Italy, is selected as the study region for implementation of the warning system. The generalized random walk- multi-state-dependent parameter method (GRW-MSDP) for soft sensor design is proposed. The GRW-MSDP estimation system is based on multi-state-dependent modeling method by utilizing the extension of the generalized random walk model. The method has been developed by utilizing the algorithms of extended Kalman filter (EKF) and fixed interval smoothing (FIS). The quality warning system of tail gas based on the estimated concentrations of $SO_2$ and $H_2S$ sends instructions to adjust the ratio of air to feed flow in the reaction furnace of SRU by plant operators. The results indicate that the proposed estimation system can be efficient in dealing with process non-linearity, high-dimensional values, and random missing data. The comparative discussion of GRW-MSDP technique performance with different soft sensing methods shows that the designed soft sensor model is more reliable with fewer input variables, lower complexity and relatively higher prediction accuracy. Furthermore, the great efficiency of the designed quality warning system is obvious from the good accuracy and F1-score values of 99.4% and 0.8951, respectively.

## Nomenclature

| | | | |
|---|---|---|---|
| $e$ | Observation error | $R$ | Covariance of one-step-ahead prediction error |
| $\mathbf{F}$ | The SDP's coefficient matrix | $\mathbf{s}$ | States that SDP is a function of them |
| $\mathbf{G}$ | The SDP's coefficient matrix | $t$ | Time numerator |
| $\mathbf{H}$ | The SDP's coefficient vector | $u$ | Input variable |

Copyright © Tavakoli (s).
Publisher: University of Sistan and Baluchestan

| | | | |
|---|---|---|---|
| $j$ | Regressor numerator | $\mathbf{X}$ | Stochastic state vector |
| $\mathbf{L}_p$ | Predecessor set | $y$ | Output variable |
| $\mathbf{L}_f$ | Successor set | $z$ | Regressor |
| $m$ | Number of delay inputs | $\alpha$ | Hyper-parameters |
| $N$ | Number of data points in time series | $\beta$ | Hyper-parameters |
| $n$ | Number of delay outputs | $\gamma$ | Hyper-parameters |
| $n_p$ | Number of parameters | $\delta$ | Hyper-parameters |
| $n_s$ | Number of state that SDP is a function of them | $\delta$ | Pure time delay |
| $\mathbf{P}$ | Covariance matrix | $\varepsilon$ | Hyper-parameters |
| $\mathbf{Q}$ | Noise variance matrix | $\varepsilon$ | One-step-ahead prediction error |
| $q$ | Polynomial order of the SDP | $\boldsymbol{\eta}$ | White noise vector with zero mean |

## I. Introduction

In this modern epoch, a significant concern to mankind is environmental pollution. Industrialization in developing and developed countries is the main reason for environmental pollution. The poor air quality in urban environments can create chronic and acute sickness in people [1]. Therefore, keeping good air quality has become an important concern for most economies. $H_2S$ and $SO_2$ are the conventional gaseous pollutants which can damage people's health and atmosphere significantly. One of the emission reasons for these gases is from industrial activities. Recently, studies have been carried out about the adequate treatment of gas containing the environmental pollutant $H_2S$ from industrial activities before discharge into the environment [2-4]. The sulfur recovery unit is a refinery processing unit which eliminates $H_2S$ from refinery gas flows before being delivered into the atmosphere. However, there are remaining $SO_2$ and $H_2S$ from the reactions in the output gas flow from the SRU (tail gas). Hence it is important to determine the tail gas quality of SRU as a safety concern issue. Estimating the tail gas quality and sending out warning signals based on the tail gas quality value is necessary to control the ratio of air to feed flow to SRU and thus avoid exposure to hazardous pollutants [5]. Many factors like the regional environmental issues and characteristics of the instruments might impress the variance of the data collected from traditional instruments, thus the costly traditional instrumentation can be replaced by soft sensors to estimate the $H_2S$ and $SO_2$ concentration. The historical data can be employed by machine learning and statistical methods to get important information that can be utilized to make decisions for safe and robust process operation. The online prediction of the product quality variables is one type of this information which can be achieved using data-driven soft sensors. The popularity of data-driven soft sensors is rising with the accessibility of stored data in industries and the accessibility of computational potency for data processing [6,7].

The first data-driven soft sensors were designed by the traditional offline modelling methods. The most popular linear offline approaches include multivariate linear regression (MLR) [8], partial least squares (PLS) [9], principal component analysis (PCA) [10] and canonical coordinates regression (CCR) [11] while the nonlinear offline approaches include support vector machine (SVM) [12] artificial neural networks (ANN) [13], neuro-fuzzy (NF) system [14] and Gaussian process regression (GPR) [15]. The nonlinear and linear offline approaches can be merged to improve the soft sensor performance such as nonlinear PLS (NNPLS) [16], nonlinear principal component regression (NNPCR) [17], kernel PLS (KPLS) [18], dynamic kernel PCA (DKPCA) [19] and least squares support vector regression (LSSVR) [20].

Catalyst activity changes, Changes of process input materials, changes in external environment and production of different grades of product quality can lead to time-varying behavior of processes thus a methodology is needed to adapt the model online [21]. Adaptive soft sensing methods are used to update the soft sensor model frequently. The moving window (MW) models [22] and the recursive updates of the PLS, least squares (LS), and PCA methods are the first adaptive soft sensors [23]. Although these methods can cope with slow process changes, they cannot adapt soft sensors with sudden changes of process. Thus, the local learning modeling methods are introduced. The popular local modeling methods include the just-in-time learning (JITL) method [24], correlation based just-in-time learning (CoJIT) method [25], locally weighted regression (LWR) [26].

If the models are adapted using faulty data, the estimation accuracy of the soft sensor reduces. Therefore, data pre-processing is employed to address data quality challenges

such as high dimensionality, inconsistencies, outliers and missing data [27]. In addition, there are methods to select contamination-free and comprehensive data from large data sets [28] and to deal with missing values such as automatically excluding the missing data and employing data imputation algorithms [29, 30].

Varying parameter models present a precise physical interpretation for the real variables of process and can be used instead of adaptive approaches [31, 32]. The models of time variable parameter (TVP) and state dependent parameter (SDP) are definitions of linear and nonlinear time-varying systems, respectively. Each parameter of univariate SDP models is varying with only one state variable and is stated utilizing the generalized random walk (GRW) model, estimated utilizing Kalman filter (KF) technique, with the implementation of fixed interval smoothing (FIS), data sorting and back-fitting algorithms [31].

The SDP-based soft sensor extension which includes the nonlinearities of multi-state dependent parameter (MSDP) is necessary for a large category of dynamic, stochastic nonlinear systems. Two best-known multi-state-dependent modeling methods are the MSDP [33] and local instrumental variable (LIV) [34] methods. In the MSDP method, the parameters are estimated using an expansion of the GRW model with the carrying out of extended FIS, data sorting and back-fitting algorithms. In the MSDP approach due to the presence of the Lagrange multipliers in the extended FIS formulation, relationships achieved by the method need to be parameterized for interpolation and dealing with the missing data. If the model parameters depend on three or more state variables, it will be difficult or impossible to parameterize the model. Also, the MSDP method cannot decrease the effect of outliers on estimation of parameters. The instrumental variable (IV) technique is employed to design the LIV method and in the LIV method, the estimate of each MSDP is not impressed by other states and regressors. But if there is a linear correlation between the model regressors, the LIV method cannot be used practically. The interpolation is performed directly without parameterization in the LIV method but the method cannot handle the missing data and cannot reduce the effect of outliers on estimated parameters. Therefore, MSDP and LIV methods need data pre-processing prior to use of the methods.

The contribution of this paper is to design and implement a new data-driven soft sensor which estimates the concentration of $SO_2$ and $H_2S$ in the tail gas flow from the SRU based on the structure of multi-state-dependent parameter (MSDP) and technique of generalized random walk. The method, termed GRW-MSDP, estimates the parameter values in the form of multivariable state space utilizing the algorithms of extended Kalman filter (KF) and fixed interval smoothing (FIS), combined with data sorting procedure and back-fitting algorithm. The extended KF decreases the effect of outlier data by assigning small weight to them and the FIS algorithm handles the missing data as a predictor-corrector type estimator. Therefore, the main advantage of GRW-MSDP method is the not-requirement of data pre-processing techniques. In the proposed method, the prediction step of the FIS algorithm is used to interpolate directly without parameterization. GRW-MSDP method can be used in the presence of linear correlation between the model regressors, due to the use of the back-fitting algorithm in accordance with the single-parameter approach. Separate from estimation, the design of a quality warning system is important to analyze the tail gas quality and to take precautionary actions to reduce the content of pollutants in the tail gas stream. Therefore, the innovation of this research is the design of a tail gas quality warning system based on $H_2S$ and $SO_2$ concentration soft sensor. In order to design the tail gas quality warning system, a classifier has been modeled based on the estimated concentration of $H_2S$ and $SO_2$.

This research mostly concentrates on the importance of the tail gas quality warning system in the SRU in the petrochemical plant of ERG PETROLI placed in Italy and the GRW-MSDP estimation system is used to achieve accurate determination of $SO_2$ and $H_2S$ concentration. Also, the SRU model prediction capability and the performance of the tail gas quality warning system are evaluated by the existence of the missing values in the industrial dataset and a comparison is conducted between the proposed approach and other soft sensing techniques.

In the rest of this article, the study region, data analysis, GRW-MSDP method and tail gas quality warning system will be given in Section II. Section III presents the results and discussions. Finally, conclusion is drawn in Section IV.

## II. Analysis and method

### *A. Study Region*

In this paper, the sulfur recovery unit in the petrochemical plant of ERG PETROLI, located in Italy, is selected as the

study region. The SRU is a refinery processing unit which recovers the elemental sulfur as a valuable byproduct from acid gases such as $H_2S$ and eliminates the $H_2S$ pollutant from

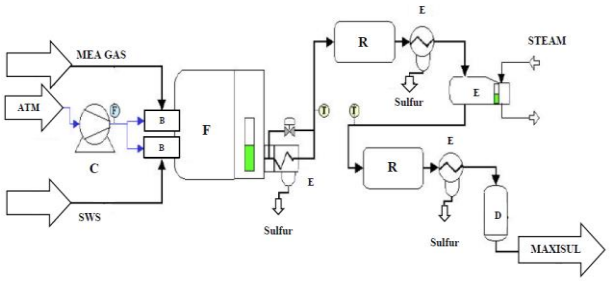


Fig. 1. Block diagram of sulfur recovery unit [5]

TABLE 1
INPUTS AND OUTPUTS OF SULFUR RECOVERY UNIT

| Variable | Description | Unit |
|---|---|---|
| $u_1$ | MEA gas flow | $m^3/hr$ |
| $u_2$ | First air flow (AIR MEA 1) | $m^3/hr$ |
| $u_3$ | Second air flow (AIR MEA 2) | $m^3/hr$ |
| $u_4$ | Air flow in SWS zone (AIR SWS) | $m^3/hr$ |
| $u_5$ | Gas flow in SWS zone | $m^3/hr$ |
| $y_1$ | $H_2S$ concentration | - |
| $y_2$ | $SO_2$ concentration | - |

gas streams. The final gas flow from the SRU (tail gas) contains remaining $SO_2$ and $H_2S$ created by reactions.

The block diagram of the SRU is demonstrated in Fig. 1. The acid gases containing $H_2S$ from the MEA plant and the acid gases containing $NH_3$ and $H_2S$ from the SWS plant are fed to a reaction furnace (F). According to the reaction (1), one third of $H_2S$ is burned in the furnace to obtain $H_2S/SO_2$ stoichiometric ratio of 2 to 1. This proportion is kept fixed during the entire process by adjusting air content in the reaction furnace.

$$H_2S + \frac{3}{2} O_2 \rightarrow SO_2 + H_2O \quad (1)$$

The exhaust gases from the furnace are sent to a water condenser. Then, in order to achieve elemental sulfur at 90% according to reaction (2), two catalytic reactors (R) and two condensers (E) are utilized.

$$2H_2S + SO_2 \rightarrow 3S + 2H_2O \quad (2)$$

Air supplies $O_2$ for the reaction (1). If the air in the reaction furnace is very little, there will be very much $H_2S$ in the tail gas. But if the air is very much, there will be very much $SO_2$ in the tail gas. Therefore, the air is an important factor in the $H_2S$ conversion. If the reactants are in stoichiometric ratio in the reaction (2), the total sulfur compounds are removed in the catalytic converters and the tail gas becomes free of pollutants. The $[H_2S]$-$2[SO_2]$ value is zero in the optimal state and is utilized as the process variable to control the ratio of air to feed flow. Thus, adjusting the air flow rate (AIR MEA 1 and AIR SWS) is manually done by the unit operators to ensure the second reaction occurs in stoichiometric proportion. The second air flow is controlled using an automatic feedback loop [5].

The concentrations of $H_2S$ and $SO_2$ residuals should be monitored due to considerable hazard for process equipment, the environment safety, and human health. Since direct measurement of their concentrations during the processing is difficult, the soft sensor methodology is suggested to estimate $SO_2$ and $H_2S$ contents.

### B. Data analysis

The soft sensor model of the sulfur recovery unit is designed based on MSDP methodology and input and output variables provided in Table 1. The benchmark dataset for the SRU contains 10081 samples, and in this paper the first 5000 samples are utilized as the train and test datasets. By the consideration of process dynamics and expert knowledge analysis, Fortuna et al. [5] presented a soft sensing model of quality parameters of SRU as follow: In order to make it easier to express, previous outputs $y_i(t\text{-}j)$ and previous inputs $u_i(t\text{-}j)$ are shown as $y_i d_j$ and $u_i d_j$ in tables and models.

$$\hat{y}_1 , \hat{y}_2 = g \{u_1, u_2, \ldots, u_5, u_1 d_5, u_1 d_7, u_1 d_9, \ldots, u_5 d_5, u_5 d_7, u_5 d_9 \} \quad (3)$$

Where $\hat{y}_1$ and $\hat{y}_2$ are the predicted values of $y_1$ and $y_2$ for the $t^{th}$ sample, respectively.

Given that the total time of reaction (2) is large and the process dynamic is slow [5], the present output value depends on the values of previous inputs. Therefore, the present inputs are eliminated and one-step delayed outputs are added in set of input variables. In order to determine the most impressive input variables, the analysis of correlation is done between the outputs and the new set of inputs using the training dataset that is shown in Table 2.

TABLE 2
CORREATION COEFFICIENTS BETWEEN OUTPUTS AT TIME t AND t-1 AGAINST INPUT DELAY VALUES

| | $u_{1d5}$ | $u_{1d7}$ | $u_{1d9}$ | $u_{2d5}$ | $u_{2d7}$ | $u_{2d9}$ |
|---|---|---|---|---|---|---|
| $y_1$ | -0.089 | -0.132 | -0.152 | -0.158 | -0.162 | -0.154 |
| $y_2$ | 0.141 | 0.175 | 0.191 | 0.202 | 0.206 | 0.198 |
| | $u_{3d5}$ | $u_{3d7}$ | $u_{3d9}$ | $u_{4d5}$ | $u_{4d7}$ | $u_{4d9}$ |
| $y_1$ | -0.083 | -0.120 | -0.139 | -0.095 | -0.096 | -0.098 |
| $y_2$ | 0.035 | 0.072 | 0.098 | 0.056 | 0.056 | 0.057 |
| | $u_{5d5}$ | $u_{5d7}$ | $u_{5d9}$ | $y_{1d1}$ | $y_{2d1}$ | |
| $y_1$ | -0.080 | -0.077 | -0.073 | 0.951 | | |
| $y_2$ | 0.043 | 0.039 | 0.036 | _ | 0.944 | |

Based on the maximum correlation coefficient in every input variable, a set of 6 impressive variables is chosen for every output variable to model the soft sensor (gray cells in Table 2). In the following, the structures of the MSDP model for output variables are given.

$$\hat{y}_1 = g \{ u_1 d_9, u_2 d_7, u_3 d_9, u_4 d_9, u_5 d_5, y_1 d_1 \} \quad (4)$$

$$\hat{y}_2 = g \{ u_1 d_9, u_2 d_7, u_3 d_9, u_4 d_9, u_5 d_5, y_2 d_1 \} \quad (5)$$

The training flowchart for soft sensor modeling based on

MSDP shown in Fig. 2 is utilized to achieve the best structure of MSDP model.

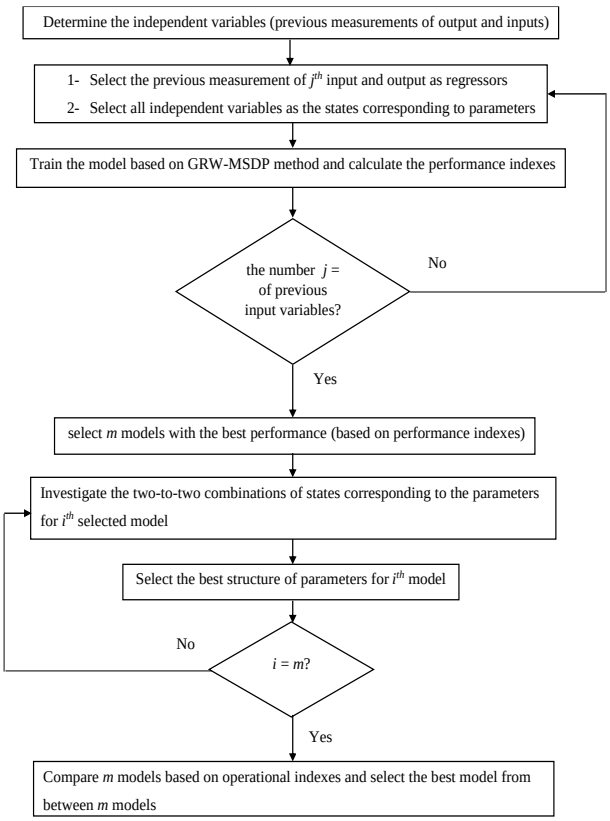


Fig. 2. Proposed flowchart for soft sensor modeling

### *C. Generalized random walk-multi-state dependent parameter method*

The generalization of state-dependent auto-regressive with exogenous variables models (SDARX) which includes the nonlinearities of multi-state-dependent parameter is defined as,

$$y_t = \sum_{j=1}^{n+m+1} x_j\left(\mathbf{s}_t^{(j)},t\right) z_{j,t} + e_t \quad \text{with} \quad \begin{cases} e_t \sim \mathrm{N}(0,\sigma^2) \\ \mathbf{s}_t^{(j)} = \left[\mathrm{s}_{1,t}^{(j)}\ \mathrm{s}_{2,t}^{(j)} \cdots \mathrm{s}_{n_{sj},t}^{(j)}\right] \end{cases} \tag{6}$$

$$z_{j,t} = \begin{cases} y_{t-j} & j \le n \\ u_{t-\delta+n+1-j} & j > n \end{cases}$$

$$x_j = \begin{cases} -\alpha_j & j \le n \\ \beta_{j-(n+1)} & j > n \end{cases}$$

$y_t$ and $u_t$ are the outputs and inputs, respectively. $z_{j,t}$ is the $j^{th}$ regressor and $n$+$m$+1 is the number of regressors in the model. $n$ and $m$ are the number of delay outputs and delay inputs, respectively. $x_j\left(\mathbf{s}_t^{(j)},t\right)$, a function of $n_{sj}$ corresponding states $(\mathrm{s}_{i,t}^{(j)}\ i = 1, 2, \ldots, n_{sj})$, is the $j^{th}$ multi-state dependent parameter, $e_t$ is a Gaussian white noise with zero mean and variance $\sigma^2$, $\delta$ is pure time delay, $\beta_{j-(n+1)}$ is the coefficient of delay ($j$-$n$-1)$^{th}$ measurement of input and $\alpha_j$ is the coefficient of delay $j^{th}$ measurement of output.

The GRW model for each SDP in a ($q_j$+1) dimensional observation space defined as the state space form in Eq. (7) is appropriate for recursive optimal estimation of multi-state-dependent parameters (MSDPs).

$$\mathbf{X}_{j,t} = \mathbf{F}_j\,\mathbf{X}_{j,t-1} + \mathbf{G}_i \tag{7}$$

$$\boldsymbol{\eta}_{j,t-1} \qquad j = 1, 2, \ldots, m+n+1 \quad , \quad t = 1, 2, \ldots, N$$

$$\mathbf{F}_j = \begin{bmatrix} \alpha & \beta \\ 0 & \gamma \end{bmatrix} \quad ; \quad \mathbf{G}_j = \begin{bmatrix} \delta & 0 \\ 0 & \varepsilon \end{bmatrix}$$

$$\mathbf{X}_{j,t} = \left[x_j\left(\mathbf{s}_t^{(j)},t\right), \nabla x_j\left(\mathbf{s}_t^{(j)},t\right), \cdots, \nabla^{q_j} x_j\left(\mathbf{s}_t^{(j)},t\right)\right]$$

where $\mathbf{X}_{j,t} = [l_{j,t},\ d_{j,t}]^{\mathrm{T}}$ is the stochastic state vector associated to $j^{th}$ state dependent parameter at $t^{th}$ sample. $d_{j,t}$ and $l_{j,t}$ are the slope change and magnitude change of correspondent SDP, respectively. $q_j$ ($q_j$ = 0, 1, 2, 3, . . .) is the polynomial order of the $j^{th}$ SDP in GRW. $\boldsymbol{\eta}_{j,t} = \left[\eta_{1j,t}\ \eta_{2j,t}\right]^T$ characterized by a covariance matrix $\mathbf{Q}_{\eta j}$ allows stochastic changes in the parameters and is a white noise vector with zero mean. The GRW model parameters are known as hyper-parameters ($\alpha$, $\beta$, $\gamma$, $\delta$, $\varepsilon$, $\mathbf{Q}_{\eta j}$). Integrated random walk (IRW) ($q_j$=1, $\alpha$=1, $\beta$=1, $\gamma$=1, $\delta$=0, $\varepsilon$=1) and random walk (RW) ($q_j$=0, $\alpha$=1, $\beta$=0, $\gamma$=0, $\delta$=1, $\varepsilon$=0) are two cases of GRW general model.

The observation equation is as follow,

$$y_t = \sum_{j=1}^{n+m+1} \mathbf{H}_{j,t}\,\mathbf{X}_{j,t} + e_t \tag{8}$$

where $\mathbf{H}_{j,t}$ represents the regressor vector $\mathbf{H}_{j,t} = [z_{j,t}\ 0]$. The formulation of state space in Eqs. (7) and (8) is appropriate to estimate the MSDPs using two passes, "forward pass filtering" and "backward pass smoothing" by the carrying out of the extended Kalman filter and recursive fixed interval smoothing (FIS), respectively. Fast changes of state variables cannot be traced by GRW model. Therefore, prior to estimation of parameters, the variables are sorted in a non-time order according to the specific state variable. In this method, Delaunay triangulation is utilized as the algorithm of data sorting. After sorting $N$ samples, the sets of predecessor ($\mathrm{L}_{p,k}$) and successor ($\mathrm{L}_{f,k}$) are defined for each $k^{th}$ sample. $n_{p,k}$ and $n_{f,k}$ are the number of the closest $p$ and $f$ samples to the $k^{th}$ sample in $\mathrm{L}_{p,k}$ and $\mathrm{L}_{f,k}$, respectively. The following numbered paragraphs display the step by step calculations of forward-pass filtering algorithm:

(1) KF algorithm includes two steps of prediction and correction. For each $k$ = 1, 2, . . . , $N$, follow the steps below.

(1-1) For each $n_{p,k}$, calculate Eqs. (9) and (10).

$$\mathbf{X}_{k,p} = \mathbf{F}\ \mathbf{X}_p \tag{9}$$

$$\mathbf{P}_{k,p} = \mathbf{F}\ \mathbf{P}_p\ \mathbf{F}^T + \mathbf{G}\,\mathbf{Q}_r\,\mathbf{G}^T \tag{10}$$

where $\mathbf{P}_{k,p}$ and $\mathbf{X}_{k,p}$ are the covariance matrix of predicted parameter vector and parameter vector prediction from every $n_{p,k}$, respectively. $\mathbf{G}$ and $\mathbf{F}$ are obtained by Eq. (7) and $\mathbf{Q}_r$ is the noise variance ratio matrix (NVR).

(1-2) Estimate the predictive parameter vector using Eqs.

(11) and (12). Then, correct the predictive parameter vector using Eqs. (13) and (14).

$$\mathbf{P}_{k|p}=\left[\frac{1}{n_{p,k}}\sum_{p\in \mathrm{L}_{p,k}}\mathbf{P}_{k,p}^{-1}\right]^{-1} \tag{11}$$

$$\mathbf{X}_{k|p}=\mathbf{P}_{k|p}\left[\frac{1}{n_{p,k}}\sum_{p\in \mathrm{L}_{p,k}}\mathbf{P}_{k,p}^{-1}\ \mathbf{X}_{k,p}\right] \tag{12}$$

$$\mathbf{P}_k=\mathbf{P}_{k|p} - \mathbf{P}_{k|p}\ \mathbf{H}_k^T\ R_k^{-1}\ \mathbf{H}_k\ \mathbf{P}_{k|p}\quad,\ R_k=1+\mathbf{H}_k\ \mathbf{P}_{k|p}\ \mathbf{H}_k^T \tag{13}$$

$$\mathbf{X}_k=\mathbf{X}_{k|p}+\mathbf{P}_k\ \mathbf{H}_k^T\ \varepsilon_k\quad,\quad \varepsilon_k=y_k-\mathbf{H}_k\ \mathbf{X}_{k|p} \tag{14}$$

$\mathbf{P}_{k|p}$ and $\mathbf{X}_{k|p}$ are the covariance matrix of predicted parameter vector and parameter vector prediction, respectively. $\mathbf{X}_k$ and $\mathbf{P}_k$ are the covariance matrix of Estimated parameter vector and parameter vector estimate, respectively. The estimate of Eqs. (11) and (12) is based on entire points that are less than *k* in sorted order and, the estimation of Eqs. (13) and (14) is based on entire points that are less than or equal to *k* in sorted order. $\varepsilon_k$ (one-step-ahead prediction error) and $R_k$ are the residual at $k^{th}$ sample and its covariance, respectively.

The EKF algorithm is just used in the process of optimizing the hyper-parameters. $\mathbf{Q}_r^{\mathrm{opt}}$ is the optimum NVR matrix and the maximum likelihood (ML) or cross validation (CV) approaches can be employed to optimize NVR hyper-parameters [33, 35]. The optimization is started with the primary values for the hyper-parameters, which either are set to default values or are chosen by the user. The concentrated likelihood and CV functions are expressed as:

$$\log(L_c)=\frac{1}{2}\left[\frac{1}{N}\sum_{k=1}^{N}\log(R_k)+\log\left(\frac{1}{N}\sum_{k=1}^{N}\frac{\varepsilon_k^2}{R_k}\right)\right] \tag{15}$$

$$\mathrm{CV}=\frac{1}{N}\sum_{k=1}^{N}{\varepsilon_k}^2 \tag{16}$$

The EKF algorithm is utilized frequently to generate $\varepsilon_k$ and $R_k$. The optimization functions are numerically minimized relative to the unknown hyper-parameters to find minimum $L_c$ and CV.

The backward pass smoothing for MSDPs estimation is carried on the following procedure:

(1) The FIS algorithm obtains a smooth estimation based on entire points in the time series and eliminates any lag effect. For each $k = N, ..., 2, 1$, estimate the smoothed parameter vector by Eqs. (17) – (20).

$$\mathbf{P}_k^*=\mathbf{Q}_P-\mathbf{Q}_P\ \mathbf{F}^T\left(\mathbf{Q}_F+\mathbf{F}\mathbf{Q}_P\mathbf{F}^T\right)^{-1}\mathbf{F}\ \mathbf{Q}_P \tag{17}$$

$$\mathbf{Q}_P=\frac{1}{n_{p,k}}\ \mathbf{G}\ \mathbf{Q}_r^{\mathrm{opt}}\ \mathbf{G}^T\quad;\quad \mathbf{Q}_F=\frac{1}{n_{f,k}}\ \mathbf{G}\ \mathbf{Q}_r^{\mathrm{opt}}\ \mathbf{G}^T$$

$$\mathbf{P}_{k|N}=\mathbf{P}_k^*-\mathbf{P}_k^*\ \mathbf{H}_k^T\left(1+\mathbf{H}_k\ \mathbf{P}_k^*\ \mathbf{H}_k^T\right)^{-1}\mathbf{H}_k\ \mathbf{P}_k^* \tag{18}$$

$$\mathbf{X}_k^*=\mathbf{F}\ \bar{\mathbf{X}}_P+\mathbf{P}_k^*\ \mathbf{F}^T\ \mathbf{Q}_F^{-1}\left(\bar{\mathbf{X}}_F-\mathbf{F}^2\ \bar{\mathbf{X}}_P\right) \tag{19}$$

$$\bar{\mathbf{X}}_P=\frac{1}{n_{p,k}}\sum_{p\ \in\ \mathrm{L}_{p,k}}\mathbf{X}_{p|N}\quad;\quad \bar{\mathbf{X}}_F=\frac{1}{n_{f,k}}\sum_{f\ \in\ \mathrm{L}_{f,k}}\mathbf{X}_{f|N}$$

$$\mathbf{X}_{k|N}=\mathbf{X}_k^*+\mathbf{P}_{k|N}\ \mathbf{H}_k^T\left(y_k-\mathbf{H}_k\ \mathbf{X}_k^*\right) \tag{20}$$

$\mathbf{P}_{k|N}$ and $\mathbf{X}_{k|N}$ are the covariance matrix of the smoothed parameter vector and the smoothed estimate of the parameter vector, respectively. Interpolation and missing data prediction are performed directly without parameterization by the prediction step ( $\mathbf{X}_k^*$ ) of FIS. The iterative Gauss-Seidel technique is used to solve the system of FIS algorithm linear equations. After estimating each MSDP, the estimated parameters need to be reordered in time order before the back-fitting algorithm carries on for the subsequent parameter.

GRW-MSDP modeling method is utilized as the core of a data-driven soft sensor for prediction of concentration of $SO_2$ and $H_2S$ in the tail gas stream.

The prediction efficiency of soft sensing models is determined using four performance indexes; determination coefficient ($R^2$), mean absolute error (MAE), root mean squares error (RMSE), and young information criterion (YIC) as follows:

$$\mathrm{R}^2=1-\frac{\sum_{j=1}^{N}\left(\hat{y}_j-\bar{y}\right)^2}{\sum_{j=1}^{N}\left(y_j-\bar{y}\right)^2} \tag{21}$$

$$\mathrm{MAE}=\frac{\sum_{j=1}^{N}\left|y_j-\hat{y}_j\right|}{N} \tag{22}$$

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{j=1}^{N}\left(y_j-\hat{y}_j\right)^2}{N}} \tag{23}$$

$$\mathrm{YIC}=\log_e^{\left\{\frac{\sigma^2}{\sigma_y^2}\right\}}+\log_e^{\mathrm{EVN}} \tag{24}$$

$$\mathrm{EVN}=\frac{1}{n_p}\sum_{j=1}^{j=n_p}\frac{\sigma^2\ p_j}{\rho_j^2}$$

$y_j$, $\hat{y}_j$ and $\bar{y}$ are the actual value, predicted value and mean values of *y*, respectively. $\rho_j$, $\sigma^2\ p_j$ and $n_p$ are parameter estimates, the estimated error variance on the $j^{th}$ parameter and the number of parameters, respectively. The YIC, a negative value, evaluates the relationship between over parameterization and model fit. The best model has the lowermost YIC, MAE and RMSE values and the highest $R^2$.

### *D. Tail gas quality warning system*

The designed tail gas quality warning system is demonstrated in Fig. 3 as a block diagram. The air and feed flow rates are fed as inputs to the designed GRW-MSDP estimation system. Then, the concentration of $H_2S$ and $SO_2$ computed by GRW-MSDP approach are given as inputs to a computational block as a classifier. The class assignment for the classifier is performed based on the criteria presented as follows, which are created according to the change in air flow rate to minimize pollutants in the tail gas stream.

Class +1: $H_2S$ concentration-2($SO_2$ concentration) > 0

Class -1: $H_2S$ concentration-2($SO_2$ concentration) < 0

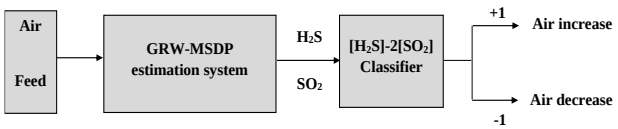


Fig. 3. Block flow diagram of tail gas quality warning system

If the classifier provides class +1 as output signal, the tail gas quality warning system produces an alarm showing that the air in the furnace is too little and tail gas contains too much of $H_2S$. Therefore, in order to adjust the $H_2S/SO_2$ stoichiometric ratio and reduce pollutants in the tail gas, the air flow must be manually increased by the unit operators to occur second reaction in the stoichiometric proportion. If the air in the furnace is too high, tail gas contains too much $SO_2$ and the classifier returns class -1 as the output. Therefore, in order to occur reaction (2) in the stoichiometric proportion and reduce pollutants in the tail gas, the air flow must be decreased manually by the operators.

The classifier performance evaluation is indicated by calculating accuracy and F1-score [36] based on the following equations.

$$\text{Accuracy} = \frac{\textit{Correct classifications}}{\textit{Number of observations}} \times 100 \tag{25}$$

$$\text{F1-score} = 2\times \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} \tag{26}$$

$$\text{Recall} = \frac{\textit{True positive}}{\textit{True positive} + \textit{False negative}}$$

$$\text{Precision} = \frac{\textit{True positive}}{\textit{True positive} + \textit{False positive}}$$

## III. Result and discussion

The model identification is performed based on RW parameter variation model and CV optimization criterion. The soft sensor model structure is obtained using the training flowchart. The training results for $y_1$ and $y_2$ are demonstrated in Table 3 and Table 4, respectively. First, all possible combinations of the regressors and states for soft sensor modeling are specified. Secondly, the models are trained based on GRW-MSDP method and the models with the best performance are selected which are shown by *. In fact, a model with a high $R^2$ and a quite low YIC is chosen, since a quite big YIC shows the possibility of over-parameterization. Then, the two-to-two combinations of states corresponding to the parameters for selected models are investigated and the best parameter structure for models are Selected (see colored cells in Table 3 and Table 4). Finally, the best configuration of the MSDP model is chosen by comparison between selected structures. Due to the use of the RW parameter variation model, when the value of the parameter NVR is very small, the parameter can be assumed to be constant. Thus, structures 6, 8, and 10 in Table 3 and Table 4 transform into 7, 9, and 11, respectively. Since the concentration of $H_2S$ and $SO_2$ in tail gas is dependent on $H_2S$ and air input to the process, the regressors and states of three mentioned structures in Table 3 are acceptable and logical in terms of expert knowledge. In Table 3, structure 11 increases the model accuracy on training data by increasing the NVR value, but it decreases the predictability of test data ($R^2_{test}$ is low). By comparing structures 7 and 9, 7 is selected as the best model structure for predicting the $H_2S$ concentration due to its lower RMSE, MAE and YIC and higher $R^2$. Since in structure 7 in Table 4, the $SO_2$ concentration is calculated by delayed amount of incoming air and one-step delayed output, this structure isn't acceptable in terms of expert knowledge. Finally, by comparing acceptable structures 9, 11 in Table 4, 11 is selected as the best model structure due to its lowest YIC, and $R^2$, RMSE, MAE and $R^2_{test}$ almost identical to the 9 structure.

The best structures of MSDP models for estimating the concentrations of $H_2S$ and $SO_2$ are determined as:

$$\hat{y}_1 = a_1\{u_3 d_9, y_1 d_1\}\ u_1 d_9 + b_1\ y_1 d_1 + e_t \tag{27}$$

$$\hat{y}_2 = a_2\{u_2 d_7, u_3 d_9\}\ u_5 d_5 + b_2\ y_2 d_1 + e_t \tag{28}$$

where $a_1$ and $a_2$ are MSDPs and $b_1$ and $b_2$ are constant parameters.

In the unfavorable conditions, the designed estimation system is a better alternative than costly traditional instruments to achieve fast and precise results and has too low RMSE of 0.01836 and 0.01549 for $H_2S$ and $SO_2$, respectively. When the missingness is not random, missing at random (MAR) happens in a dataset, and in this case the missingness is totally accounted for by variables that contain complete information. To demonstrate the robustness of the proposed soft sensor modeling method, the output value of 550 random samples from the training dataset of $y_1$ and $y_2$ are missing according to the MAR mechanism. Figs. 4 and 5 show the real and predicted values of $y_1$ and $y_2$ on the training dataset, respectively. The missing values are indicated by regions marked by rectangles. According to the results, the estimated and real values agree well with each other in both $y_1$ and $y_2$ and the used method can accurately predict the missing values.

Among the total of 2500 data samples utilized for classifier in the presence of missing values, there are 17 misclassifications and 2483 data samples are classified correctly. In the 17 misclassifications, 10 are false negatives and 7 are false positives. In the 2483 correct classifications, there are 2419 true negatives and 64 true positives. Accuracy and F1-score are obtained as 99.32% and 0.8828, respectively for the tail gas quality warning system demonstrating that the designed quality warning system is reliable in the presence of missing values.

TABLE 3

TRAINING RESULTS FOR $y_1$ WITH VARIOUS REGRESSORS AND STATES

| | regressors | states | NVR | $R^2$ | RMSE | MAE | YIC | $R^2_{test}$ |
|---|---|---|---|---|---|---|---|---|
| 1* | $u_1d_9$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 4.95305 | 0.90608 | 0.02053 | 0.00582 | -63.1956 | 0.88536 |
| | $y_1d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 1.20519e-18 | | | | | |
| 2 | $u_2d_7$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 1.93132e+1 | 0.90818 | 0.02029 | 0.00580 | -53.7448 | 0.88350 |
| | $y_1d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 1.63352e-14 | | | | | |
| 3* | $u_3d_9$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 3.77913 | 0.90480 | 0.02068 | 0.00628 | -58.6019 | 0.88291 |
| | $y_1d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 1.18134e-16 | | | | | |
| 4 | $u_4d_9$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 2.38604e+1 | 0.90975 | 0.02015 | 0.00585 | -54.8812 | 0.88479 |
| | $y_1d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 5.32083e-15 | | | | | |
| 5* | $u_5d_5$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 1.58471 | 0.90512 | 0.02067 | 0.00604 | -64.6855 | 0.88552 |
| | $y_1d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_1d_1$ | 2.65594e-19 | | | | | |
| 6 | $u_1d_9$ | $(u_3d_9, y_1d_1)$ | 2.13511 | 0.92297 | 0.01859 | 0.00519 | -66.4423 | 0.87593 |
| | $y_1d_1$ | $(u_4d_9, u_5d_5)$ | 2.86497e-21 | | | | | |
| 7 | $u_1d_9$ | $(u_3d_9, y_1d_1)$ | 1.80000 | 0.92477 | 0.01836 | 0.00502 | -20.9464 | 0.88787 |
| | $y_1d_1$ | (-) | 0 | | | | | |
| 8 | $u_3d_9$ | $(u_1d_9, u_2d_7)$ | 8.12757e-1 | 0.91196 | 0.01987 | 0.00560 | -57.7681 | 0.88787 |
| | $y_1d_1$ | $(u_1d_9, u_3d_9)$ | 1.77828e-17 | | | | | |
| 9 | $u_3d_9$ | $(u_1d_9, u_2d_7)$ | 3.14381e-1 | 0.91077 | 0.02002 | 0.00559 | -20.6170 | 0.89052 |
| | $y_1d_1$ | (-) | 0 | | | | | |
| 10 | $u_5d_5$ | $(u_2d_7, u_3d_9)$ | 3.82804 | 0.94626 | 0.01552 | 0.00458 | -53.4242 | 0.89056 |
| | $y_1d_1$ | $(u_1d_9, y_1d_1)$ | 3.13960e-15 | | | | | |
| 11 | $u_5d_5$ | $(u_2d_7, u_3d_9)$ | 3.48962 | 0.94654 | 0.01548 | 0.00475 | -21.4786 | 0.86515 |

TABLE 4

TRAINING RESULTS FOR $y_2$ WITH VARIOUS REGRESSORS AND STATES

| | regressors | states | NVR | $R^2$ | RMSE | MAE | YIC | $R^2_{test}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | $u_1d_9$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 3.29193 | 0.99342 | 0.00447 | 0.00169 | -19.8493 | 0.88826 |
| | $y_2d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 1.21432e+4 | | | | | |
| 2 | $u_2d_7$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 1.11852e+1 | 0.98359 | 0.00706 | 0.00283 | -18.1709 | 0.88801 |
| | $y_2d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 4.88249e+3 | | | | | |
| 3* | $u_3d_9$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 5.47454 | 0.99704 | 0.00300 | 0.00104 | -21.3284 | 0.88868 |
| | $y_2d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 2.43154e+4 | | | | | |
| 4* | $u_4d_9$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 1.61668 | 0.99743 | 0.00279 | 0.00093 | -21.5301 | 0.88617 |
| | $y_2d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 2.94792e+4 | | | | | |
| 5* | $u_5d_5$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 1.06796 | 0.99599 | 0.00349 | 0.00122 | -20.7206 | 0.88684 |
| | $y_2d_1$ | $u_1d_9, u_2d_7, u_3d_9, u_4d_9, u_5d_5, y_2d_1$ | 1.97858e+4 | | | | | |
| 6 | $u_3d_9$ | $(u_4d_9, u_5d_5)$ | 8.53372e-19 | 0.96103 | 0.01088 | 0.00525 | -55.2530 | 0.88752 |
| | $y_2d_1$ | $(u_2d_7, u_4d_9)$ | 5.98381e+1 | | | | | |
| 7 | $u_3d_9$ | (-) | 0 | 0.95882 | 0.01118 | 0.00528 | -18.0488 | 0.88722 |
| | $y_2d_1$ | $(u_2d_7, u_4d_9)$ | 5.30959e+1 | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | $u_4d_9$ | $(u_1d_9, u_5d_5)$ | 3.86188e-21 | 0.92065 | 0.01552 | 0.00652 | -57.0638 | 0.88715 |
| | $y_2d_1$ | $(u_1d_9, u_5d_5)$ | 4.61123 | | | | | |
| 9 | $u_4d_9$ | (-) | 0 | 0.92165 | 0.01542 | 0.00658 | -18.2832 | 0.88716 |
| | $y_2d_1$ | $(u_1d_9, u_5d_5)$ | 4.86968 | | | | | |
| 10 | $u_5d_5$ | $(u_2d_7, u_3d_9)$ | 1.20091 | 0.91762 | 0.01581 | 0.00640 | -56.5420 | 0.88809 |
| | $y_2d_1$ | $(u_4d_9, y_2d_1)$ | 9.29781e-17 | | | | | |
| 11 | $u_5d_5$ | $(u_2d_7, u_3d_9)$ | 1.43076 | 0.92094 | 0.01549 | 0.00640 | -22.2774 | 0.88739 |
| | $y_2d_1$ | (-) | 0 | | | | | |

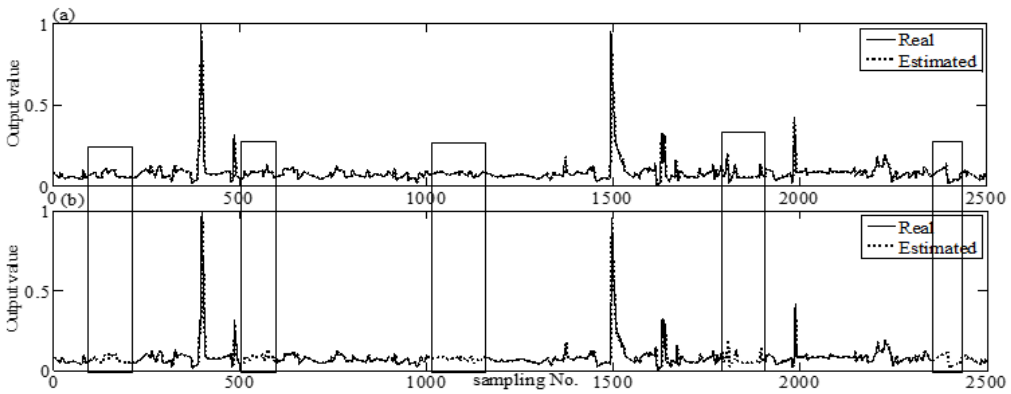


Fig. 4. $H_2S$ concentration prediction results on train dataset (a) free of missing data, (b) with missing data

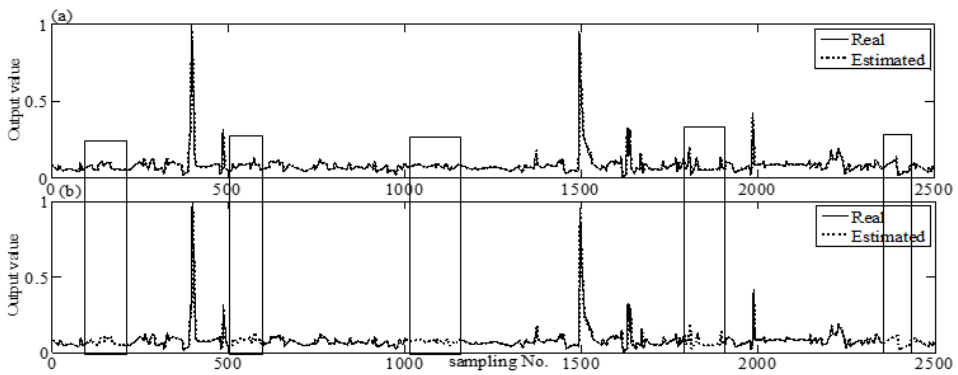


Fig. 5. $SO_2$ concentration prediction results on train dataset (a) free of missing data, (b) with missing data

Figs. 6(a) and 6(b) display the real and estimated values of $y_1$ and $y_2$ on the testing dataset. Since the estimated and real values agree well with each other in both $y_1$ and $y_2$, the GRW-MSDP method can properly model the nonlinearity. Among a total of 2500 samples utilized for the classifier in the testing dataset, there are 15 misclassifications and 2485 data samples are classified correctly. In the 15 misclassifications, 10 are false negatives and 5 are false positives. In the 2485 correct classifications, there are 2421 true negatives and 64 true positives. Accuracy and F1-score are obtained as 99.4% and 0.8951, respectively for the tail gas quality warning system demonstrating that the designed warning system is reliable in testing dataset.

The scatter plots of output concentration estimations against real values are presented in Figs. 7(a) and 7(b). The estimated and real values are in good compromise with each other and most of the data samples are situated along the diameter line, which displays great prediction accuracy of the soft sensor.

The average relative error between the predicted and actual output values of $y_1$ and $y_2$ for the whole set of 5000 samples are 5.54% and 6.28%, respectively, which show the proper performance of soft sensor prediction. The GRW-MSDP method prediction performance is compared with different soft sensing approaches for the SRU in Table 5.

According to the performance indexes presented in Table 5, linear models such as MLR, PCR and PLS have a relatively high RMSE and have displayed undesirable performance in dealing with the nonlinear data. The RMSE of several models such as LASS, OLPLS, SELPLS, ELFIR-SKRLS, MWtrPBC, MWAdp –JITL, $JITL_{TT}$ –MWt, SDP, MSDP, C-vine CR, D-vine CR, R-vine CR, JITL-GRNN and SCN are too close to the GRW-MSDP approach. In the mentioned models (With the exception of SDP, MSDP, MWAdp –JITL, C-vine CR, D-vine CR, R-vine CR and JITL-GRNN), the output variables are

predicted using 20 or more input variables, while more input variables lead to high model complexity and high computational load.

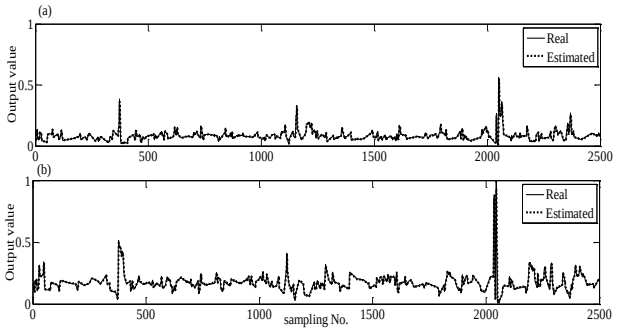


Fig. 6. Estimated output results on test dataset (a) $y_1$, (b) $y_2$

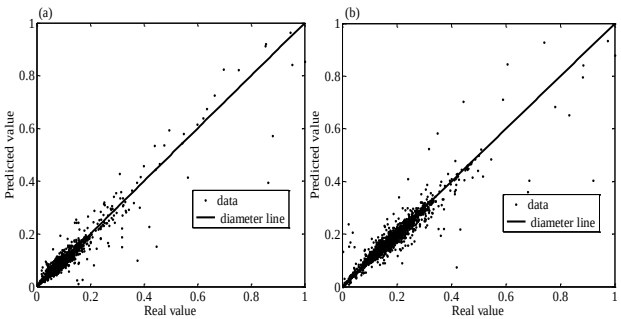


Fig. 7. Estimated values against real values of output concentration (a) $y_1$, (b) $y_2$

In MWAdp –JITL, C-vine CR, D-vine CR, R-vine CR and JITL-GRNN, the input delay values were not considered as input variables, so the designed soft sensor models are not appropriate for control aims. None of the mentioned soft sensors except SDP and MSDP were considered the system dynamic (delayed outputs), while if the previous information of the output variable is considered, it has a significant effect on improving the quality prediction with fewer input variables. The current modeling method can be included in the category of the best methods due to few inputs, simple model, low computational load, suitable performance indicators, direct interpolation without parameterization and automatic dealing with missing data.

## IV. Conclusions

$H_2S$ and $SO_2$ are dangerous air pollutants and their estimation is important for keeping environmental health. The tail gas stream from the industrial sulfur recovery unit contains remaining $SO_2$ and $H_2S$ created by reactions and it is significant to design an efficient tail gas quality warning system. The SRU in the petrochemical plant of ERG PETROLI in Italy is selected as the study region for implementation of the warning system. A machine learning approach namely GRW-MSDP is suggested for the development of an efficient soft sensor to compute the concentration of $SO_2$ and $H_2S$. The performance evaluation of the designed estimation system based on GRW-MSDP demonstrates that the suggested method is more efficient in modeling nonlinear systems and in dealing with random missing values, and high-dimensional data. A comparative study between different estimation methods for SRU has shown that the proposed method considerably reduces the error while is more robust and reliable with fewer input variables, less complexity, and relatively higher prediction accuracy over other presented estimation approaches. A fast classifier is utilized to design the tail gas quality warning system based on the suggested soft sensor and the accuracy and F1-score of the warning system are obtained as 99.4% and 0.8951, respectively. The precise results for GRW-MSDP estimate system and classifier indicate that the designed tail gas quality warning system is reliable.

TABLE 5
COMPARSION OF PREDICTION RESULTS OF VARIOUS SOFT SENSORS FOR SRU

| Publication | Model type | No. of variables | $H_2S$ concentration | | | $SO_2$ concentration | | |
|---|---|---|---|---|---|---|---|---|
| | | | $R^2$ | RMSE | MAE | $R^2$ | RMSE | MAE |
| Fortuna et al. [5] | MLP[1] | 20 | - | 0.0300 | - | - | 0.0200 | - |
| | RBF[2] | 20 | - | 0.0424 | - | - | 0.0387 | - |
| | NF | 20 | - | 0.0346 | - | - | 0.0283 | - |
| | Nonlinear LSQ[3] | 20 | - | 0.0283 | - | - | 0.0200 | - |
| Ge and Song [37] | PLS | 5 | - | 0.0528 | - | - | 0.0563 | - |
| | LSSVM[4] | 5 | - | 0.0401 | - | - | 0.0512 | - |
| | RVM[5] | 5 | - | 0.0397 | - | - | 0.0511 | - |
| Shao et al. [38] | JITPLS | 20 | - | 0.0210 | - | - | 0.0238 | - |
| | RPLS[6] | 20 | - | 0.0196 | - | - | 0.0209 | - |
| | OLPLS[7] | 20 | - | 0.0162 | - | - | 0.0142 | - |
| Shao and Tian | LWPLS[8] | 20 | - | 0.0181 | - | - | 0.0179 | - |

[1] Multi-Layer Perceptrons
[2] Radial Basis Function
[3] Nonlinear Least Square
[4] Least-Squares Support Vector Machine
[5] Relevance Vector Machine
[6] Recursive PLS
[7] Online Local PLS
[8] Locally Weighted PLS

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| [39] | MWPLS[9] | 20 | - | 0.0246 | - | - | 0.0230 | - |
| | LASS[10] | 20 | - | 0.0172 | - | - | 0.0168 | - |
| | SELPLS[11] | 20 | - | 0.0130 | - | - | 0.0119 | - |
| Jain et al. [40] | SSVR[12] | 5 | - | 0.068 | 0.032 | - | 0.067 | 0.045 |
| | LSSVR[13] | 5 | - | 0.078 | 0.051 | - | 0.123 | 0.094 |
| | MLR | 5 | - | 0.069 | 0.037 | - | 0.07 | 0.048 |
| | PCR | 5 | - | 0.060 | 0.037 | - | 0.051 | 0.048 |
| Shao and Tian [41] | PLS | 20 | - | 0.0519 | - | - | 0.0573 | - |
| | LSSVR | 20 | - | 0.0415 | - | - | 0.0517 | - |
| | JITL LSSVR | 20 | - | 0.0401 | - | - | 0.0475 | - |
| | EnLSSVR[14] | 20 | - | 0.0382 | - | - | 0.0477 | - |
| | SEnL-DM[15] | 20 | - | 0.0376 | - | - | 0.0421 | - |
| | $S^3$EnL-DM[16] | 20 | - | 0.0367 | - | - | 0.0408 | - |
| Bidar et al. [33] | SDP | 4 | - | 0.0139 | 0.0055 | - | 0.0144 | 0.0073 |
| | MSDP | 3 and 4 | - | 0.0119 | 0.0034 | - | 0.0121 | 0.0065 |
| Chen et al. [42] | ELFIR-LS[17] ($N^S$=5) | 150 | - | 0.0216 | - | - | - | - |
| | ELFIR-PLS ($N^S$=5) | 150 | - | 0.0219 | - | - | - | - |
| | ELPLS ($N^S$=5) | 150 | - | 0.02125 | - | - | - | - |
| | EL*FIR-SKRLS[18] ($N^S$=5) | 150 | - | 0.0215 | - | - | - | - |
| | ELFIR-SKRLS ($N^S$=5) | 150 | - | 0.01975 | - | - | - | - |
| Alakent [43] | $MW_{tr}$ [PBC] | 30 | 0.910 | 0.0153 | - | 0.931 | 0.0135 | - |
| Urhan and Alakent [44] | $MW_{Adp}$ -JITL | 5 | 0.932 | 0.0136 | 0.0081 | - | - | - |
| Alakent [45] | $JITL_{TT}$ -$MW_tr$ | 30 | 0.914 | 0.0159 | - | 0.934 | 0.0138 | - |
| Liu et al. [46] | C-vine CR[19] | 5 | - | - | - | 0.9556 | 0.0101 | - |
| | D-vine CR | 5 | - | - | - | 0.9519 | 0.0106 | - |
| | R-vine CR | 5 | - | - | - | 0.9625 | 0.0093 | - |
| Vijayan et al.[47] | JITL-GRNN[20] | 7 | - | 0.0180 | 0.0064 | - | 0.0165 | 0.0080 |
| Zhang et al. [48] | CDADPLS[21] | 5 | - | 0.5855 | - | - | 0.3821 | - |
| Zhao et al. [49] | SCN[22] | 20 | - | - | - | 0.9524 | 0.0105 | 0.0077 |
| | CTSCN[23] | 20 | - | - | - | 0.8396 | 0.0224 | 0.0164 |
| | LPSCN[24] | 20 | - | - | - | 0.8408 | 0.0205 | 0.0161 |
| This work | GRW-MSDP | 4 | 0.92477 | 0.01836 | 0.00502 | 0.92094 | 0.01549 | 0.00640 |

## References


[1] H. Xu, Y. Jia, Z. Sun, J. Su, Q.S. Liu, Q. Zhou, and G. Jiang, "Environmental pollution, a hidden culprit for health issues, " Eco-Environment & Health, Vol.1, No.1, pp.31-45, 2022.

[2] Z. Zhao, J. Yang, Z. Zhang, S.Wang, Z. Zhang, and J. Lu, "New method for efficient control of hydrogen sulfide and methane in gravity sewers: Combination of NaOH and Nitrite, " Front. Environ. Sci. Eng., Vol. 16, No. 6, pp. 1-14, 2021.

[3] Y. Wang, P. Ning, R. Zhao, K. Li, C. Wang, X. Sun, X. Song, and Q. Lin, "A Cu-modified active carbon fiber significantly promoted H2S and PH3 simultaneous removal at a low reaction temperature, " Front. Environ. Sci. Eng., "Vol. 15, No. 6, pp. 1-10, 2021.

[4] S. Sun, T. Jia, K. Chen, Y. Peng, and L. Zhang, "Simultaneous removal of hydrogen sulfide and volatile organic sulfur compounds in off-gas mixture from a wastewater treatment plant using a two-stage bio-trickling filter system, " Front. Environ. Sci. Eng., Vol. 13, No. 4, pp. 1-13, 2019.

[5] L. Fortuna, A. Rizzo, M. Sinatra, and M.G. Xibilia, "Soft analyzers for a sulfur recovery unit, " Control Engineering Practice, Vol. 11, No. 12, pp. 1491-1500, 2003.

[6] Y. Yu, M. Peng, H. Wang, Z. Ma, and W. Li, "Improved PCA model for multiple fault detection, isolation and reconstruction of sensors in nuclear power plant, " Annals of Nuclear Energy, Vol. 148, p. 107662, 2020.

[7] Y.S. Perera, D.A.A.C. Ratnaweera, C.H. Dasanayaka, and C. Abeykoon, "The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review, " Engineering Applications of Artificial Intelligence, Vol. 121, p.105988, 2023.

[8] G.K. Uyanık and N. Güler, "A study on multiple linear regression analysis, " Procedia-Social and Behavioral


[9] Moving Window PLS
[10] Localized Adaptive Soft Sensor
[11] Selective Ensemble of Local PLS
[12] Smooth SVR
[13] Least Square SVR
[14] Ensemble of SSVR
[15] Supervised Ensemble Learning Based on Distance to Model
[16] Selective Semi-Supervised EnL-DM
[17] Ensemble Local Global Finite Impulse Response-LS
[18] Stable Kernel-based Regularized LS
[19] Vine Copula Regression
[20] JITL-Generalized Regression Neural Network
[21] Concept Drift Adaptive Dynamic PLS
[22] Stochastic Configuration Network
[23] Co-Training SCN
[24] Locality Preserving SCN

Sciences, Vol. 106, pp. 234-240, 2013.
[9] Q. Zhu, "Dynamic Autoregressive Partial Least Squares for Supervised Modeling, " IFAC-PapersOnLine, Vol, 54, No. 7, pp. 234-239, 2021.
[10] D. Yan, F.Yang, F.Yang, H. Zhang, Z. Guo, J. Li, and Y. Wu, "Identifying the key system parameters of the organic Rankine cycle using the principal component analysis based on an experimental database, " Energy Conversion and Management, Vol. 240, p. 114252, 2021.
[11] A.J. Burnham, R. Viveros, and J.F. MacGregor, "Frameworks for latent variable multivariate regression, " Journal of chemometrics, Vol. 10, No. 1, pp. 31-45, 1996.
[12] J. Huang, T. Jin, M. Liang, and H. Chen, "Prediction of heat exchanger performance in cryogenic oscillating flow conditions by support vector machine, " Applied Thermal Engineering, Vol. 182, p. 116053, 2021.
[13] S. Kamat and K. Madhavan, "Developing ANN based virtual/soft sensors for industrial problems, " IFAC-PapersOnLine, Vol. 49, No. 1, pp. 100-105, 2016.
[14] P.V. de Campos Souz, "Fuzzy neural networks and neuro-fuzzy networks: A review the main techniques and applications used in the literature, " Applied soft computing, Vol. 92, p. 106275, 2020.
[15] J. Chen, J. Yu, and Y. Zhang, "Multivariate video analysis and Gaussian process regression model based soft sensor for online estimation and prediction of nickel pellet size distributions, " Computers & Chemical Engineering, Vol. 64, pp. 13-23, 2014.
[16] E. Zamprogna, M. Barolo, and D.E. Seborg, "Development of a soft sensor for a batch distillation column using linear and nonlinear PLS regression techniques, " IFAC Proceedings Volumes, Vol. 35, No. 1, pp. 431-436, 2002.
[17] Q.X. Zhu, X. Wang, Y.L. He, and Y. Xu, "An improved extreme learning machine integrated with nonlinear principal components and its application to modeling complex chemical processes, " Applied Thermal Engineering, Vol. 130, pp. 745-753, 2018.
[18] X. Zhang, W. Yan, and H. Shao, "Nonlinear multivariate quality estimation and prediction based on kernel partial least squares, " Industrial & engineering chemistry research, Vol. 47, No. 4, pp. 1120-1131, 2008.
[19] S.W. Choi and I.B. Lee, "Nonlinear dynamic process monitoring based on dynamic kernel PCA, " Chemical engineering science, Vol. 59, No. 24, pp. 5897-5908, 2004.
[20] W. Zhou, Y. Cheng, S. Ding, L. Chen, and R. Li, "A grey seasonal least square support vector regression model for time series forecasting, " ISA transactions, Vol. 114, pp. 82-98, 2021.
[21] P. Kadlec, R. Grbić, and B. Gabrys, "Review of adaptation mechanisms for data-driven soft sensors, " Computers & chemical engineering, Vol. 35, No. 1, pp.1-24, 2011.
[22] Y. Liu, D. Huang, B. Liu, Q. Feng, and B. Cai, "Adaptive ranking based ensemble learning of Gaussian process regression models for quality-related variable prediction in process industries, " Applied Soft Computing, Vol. 101, p. 107060, 2021.
[23] I. Portnoy, K. Melendez, H. Pinzon, and M. Sanjuan, "An improved weighted recursive PCA algorithm for adaptive fault detection, " Control Engineering Practice, Vol. 50, pp. 69-83, 2016.
[24] W.S. Yeo, A. Saptoro, P. Kumar, and M. Kano, "Just-in-time based soft sensors for process industries: A status report and recommendations, " Journal of Process Control, Vol. 128, p.103025, 2023.
[25] K.Fujiwara, M. Kano, S. Hasebe, and A. Takinami, "Soft - sensor development using correlation - based just - in - time modeling, " AIChE Journal, No. 55, Vol. 7, pp. 1754-1765, 2009.
[26] K. Liu, W. Shao, and G. Chen, "Autoencoder-based nonlinear Bayesian locally weighted regression for soft sensor development, " ISA transactions, Vol. 103, pp. 143-155, 2020.
[27] M. Frye, J. Mohren, and RH. Schmitt, "Benchmarking of Data Preprocessing Methods for Machine Learning-Applications in Production, " Procedia CIRP, Vol. 104, pp. 50-55, 2021.
[28] S. Cohen, Dealing with data: strategies of preprocessing data. In: Artificial Intelligence and Deep Learning in Pathology, Amsterdam, Netherlands: Elsevier, 2021, pp. 77-92.
[29] Huber P.J., " Robust statistics " volume 523, John Wiley & Sons: Hoboken, New Jersey, United States (2004).
[30] L. Ma, M. Wang, and K. Peng, "A missing manufacturing process data imputation framework for nonlinear dynamic soft sensor modeling and its application," Expert Systems with Applications, Vol. 237, p. 121428, 2024.
[31] B. Bidar, J. Sadeghi, F. Shahraki, and M.M. Khalilipor, "Data-driven soft sensor approach for online quality prediction using state dependent parameter models, " Chemometrics and Intelligent Laboratory Systems, Vol. 162, pp. 130-141, 2017.
[32] R. Parvizi Moghadam, J. Sadeghi, and F. Shahraki, "Optimization of time - variable - parameter model for data - based soft sensor of industrial debutanizer, " Optimal Control Applications and Methods, Vol. 41, No.2, pp. 381-394, 2020.
[33] B. Bidar, F. Shahraki, J. Sadeghi, and M.M. Khalilipor, "Soft sensor modeling based on multi-state-dependent parameter models and application for quality monitoring in industrial sulfur recovery process, " IEEE Sensors Journal, Vol. 18, No. 11, pp. 4583-4591, 2018.
[34] B. Bidar, M.M. Khalilipor, F. Shahraki, and J. Sadeghi, "A data-driven soft-sensor for monitoring ASTM-D86 of CDU side products using local instrumental variable (LIV) technique, " Journal of the Taiwan Institute of Chemical Engineers, Vol. 84, pp. 49-59, 2018.
[35] DM. Hawkins, S.C. Basak, and D. Mills, "Assessing model fit by cross-validation, " Journal of chemical information and computer sciences, Vol. 43, No. 2, pp.579-586, 2003.
[36] B. Sprague, Q. Shi, M.T. Kim, L. Zhang, A. Sedykh, E. Ichiishi, H. Tokuda, K.H. Lee, and H. Zhu, "Design, synthesis and experimental validation of novel potential chemopreventive agents using random forest and support vector machine binary classifiers, " Journal of computer-aided molecular design, Vol. 28, No. 6, pp. 631-646, 2014.
[37] Z. Ge and Z. Song, "Nonlinear soft sensor development based on relevance vector machine, " Industrial & engineering chemistry research, Vol. 49, No. 18, pp. 8685-8693, 2010.
[38] W. Shao, X. Tian, and P. Wang, "Local partial least squares based online soft sensing method for multi-output processes with adaptive process states division, " Chinese Journal of Chemical Engineering, Vol. 22, No. 7, pp. 828-836, 2014.
[39] W. Shao and X. Tian, "Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models, " Chemical Engineering Research and Design, Vol. 95, pp.113-132, 2015.
[40] V. Jain, P. Kishore, R.A. Kumar, and A.K. Pani, "Inferential sensing of output quality in petroleum refinery using principal component regression and support vector

regression, " In: 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, January 5-7, 2017, Piscataway: IEEE, 2017, pp. 461-465.
[41] W. Shao and X. Tian, "Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development, " Neurocomputing, Vol. 222, pp. 91-104, 2017.
[42] X. Chen, Z. Mao, R. Jia, and S. Zhang, "Ensemble regularized local finite impulse response models and soft sensor application in nonlinear dynamic industrial processes, " Applied Soft Computing, Vol. 85, p. 105806, 2019.
[43] B. Alakent, "Soft sensor design using transductive moving window learner, " Computers & Chemical Engineering, Vol. 140, p.106941, 2020.
[44] A. Urhan and B. Alakent, "Integrating adaptive moving window and just-in-time learning paradigms for soft-sensor design, " Neurocomputing, Vol. 392, pp. 23-37, 2020.
[45] B. Alakent, "Soft-sensor design via task transferred just-in-time-learning coupled transductive moving window learner, " Journal of Process Control, Vol. 101, pp. 52-67, 2021.
[46] S. Liu, and S. Li, "A semi-supervised soft sensor method based on vine copula regression and tri-training algorithm for complex chemical processes," Journal of Process Control, Vol. 120, pp.115-128, 2022.
[47] V. Vijayan, H.K. Mohanta, and A.K. Pani, " Adaptive non-linear soft sensor for quality monitoring in refineries using Just-in-Time Learning—Generalized regression neural network approach," Applied Soft Computing, Vol. 119, p.108546, 2022.
[48] T. Zhang, G. Yan, M. Ren, L. Cheng, R. Li, and G. Xie, " Dynamic transfer soft sensor for concept drift adaptation," Journal of Process Control, Vol. 123, pp.50-63, 2023.
[49] Y. Zhao, X. Deng, and S. Li, " A nonlinear industrial soft sensor modeling method based on locality preserving stochastic configuration network with utilizing unlabeled samples, " ISA transactions, 2023.

**Fereshte Tavakoli Dastjerd** received the B.S., M.S. and Ph.D. degrees in chemical engineering from University of Sistan and Baluchestan, Zahedan, Iran in 2013, 2016 and 2022, respectively. Her research interests include process modeling and simulation, process identification, process control and soft sensor design for chemical processes.

**Farhad Shahraki** received his B.S. degree in chemical engineering from the University of Sistan and Baluchestan, Zahedan, Iran in 1989, his M.S. degree in chemical engineering from Tarbiat Modares University, Tehran, Iran in 1992, and his Ph.D. degree in chemical engineering from The University of Manchester, Manchester, the U.K. in 2001. He is currently a professor of chemical engineering at the University of Sistan and Baluchestan. His research interests include process integration, process modeling, and optimization.

**Jafar Sadeghi** received his B.S. in chemical engineering from the Isfahan University of Technology, Isfahan, Iran, his M.S. degree in chemical engineering from the Sharif University of Technology, Tehran, Iran and his Ph.D. degree in chemical engineering from University of Lancaster, the U.K. in 1991, 1995, and 2007, respectively. He is currently an associate professor at the University of Sistan and Baluchestan, Zahedan, Iran. His research interests cover process modeling and simulation, process control, process identification, process intensification, automation, and instrumentation.

**Mir Mohammad Khalilipour** received his bachelor's degree in chemical engineering from Quchan University in 2004 and his master's and Ph.D. degrees in chemical engineering process design from the University of Sistan and Baluchistan, Iran in 2008 and 2016, respectively. He has been involved in safety management industrial projects for many Iranian oil and gas refineries from 2008 to 2011. He is currently an assistant professor at the Chemical Engineering Department, the University of Sistan and Baluchestan. His research interests are chemical safety, control, and process simulation especially with a focus on practical interesting of control, and soft sensors in chemical processes.

**Bahareh Bidar** received her bachelor's degree in chemical engineering from the Ferdowsi University of Mashhad, Iran, in 2007 and her master's and Ph.D. degrees in chemical engineering process design from the University of Sistan and Baluchistan, Iran in 2012 and 2018, respectively. Her main areas of research interest are process modeling and simulation, especially in soft sensor design for chemical process, process identification and control. She is currently an assistant professor in the chemical engineering department at the University of Sistan and Baluchestan.