IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 November 2022, accepted 4 December 2022, date of publication 9 December 2022,
date of current version 14 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228048

RESEARCH ARTICLE

A Multi-Rate Probabilistic Slow Feature Regression Model for Dynamic Feature Learning and Industrial Soft Sensor Development

MIAO ZHANG[1], ZHIWEI WEN[2], AND LE ZHOU[1], (Member, IEEE)
[1]School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China
[2]Zhejiang Huashi Technology Company Ltd., Hangzhou 310023, China

Corresponding authors: Miao Zhang (zhangmiao@zust.edu.cn) and Le Zhou (zhoule@zust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62003306, and in part by the Natural Science Foundation of Zhejiang Province under Grant LQ21F030009.

ABSTRACT In practical process industries, the measurements coming from different sources are collected at different sampling rates, thereby soft sensors developed using uniformly sampled measurements may result in poor prediction performance. Besides, industrial processes are inherently stochastic and most of them present dynamic characteristic. To cope with these issues, a multi-rate probabilistic slow feature regression (MR-PSFR) model is proposed in this paper for dynamic feature learning and soft sensor development in industrial processes. In the MR-PSFR, both input and output observation datasets with different sampling rates are used to extract the slow features, which can separate slowly and fast changing features and have a better interpretation of the outputs. Then, the expectation-maximization algorithm is modified to derive the model parameters of MR-PSFR and the quality prediction strategy for multi-rate processes is constructed. Finally, the proposed method is investigated through a numerical example and a real industrial process. The simulation results show that the extracted slow features better represent the intrinsic characteristics of the processes and the proposed model has better prediction performance for multi-rate dynamic processes than other methods.

INDEX TERMS Soft sensor, feature learning, multi-rate probabilistic slow feature regression, expectation-maximization algorithm, multi-rate dynamic process.

I. INTRODUCTION

In industrial processes, accurate measurement of quality variables is of great significance for process control, monitoring and optimization. At present, quality variables are mainly measured by online sampling and off-line analysis in the laboratory. The measurement process is time-consuming and the related detection equipment is expensive, so it is not conducive to realize the real-time monitoring and control of the industrial process [1], [2], [3], [4]. As a supplement of traditional measurement methods, soft sensor technology can solve this problem. Data-driven soft sensors can estimate quality variables accurately and economically by establishing mathematical models between key quality variables and easily measurable auxiliary variables [5], [6], [7]. With the wide application of distributed control system (DCS), a large amount of process data reflecting the real process state has been collected and stored. On this basis, various data-driven soft sensors have been developed and successfully applied in practical industrial processes [8], [9], [10], [11].

However, most traditional soft sensor methods assume a consistent sampling rate for all the observations. In most chemical processes, the sampling rates of the process and quality variables may vary among a large range. Those important quality variables are tested at the laboratory with a much lower sampling rate, such as the melt index of polypropylene, the content of butane, the endpoint of crude oil. They may be collected among a few hours or days. On the other hand, basic

The associate editor coordinating the review of this manuscript and approving it for publication was Qingchao Jiang.

process variables such as pressure, flow rate and temperature, can be measured online using high-rate sensors, resulting in the multi-rate characteristic of the industrial process [12], [13]. The data imbalance between the quality variables and the process variables makes it challenging to build an accurate estimation model for quality variables.

For quality prediction, the traditional method is to convert the multi-rate sampling data into single-rate sampling data. Usually, the data preprocessing techniques include up-sampling and down-sampling. The up-sampling method uses high sampling rate data to estimate the uncollected data with low sampling rate by establishing a regression model. In the down-sampling method, all variables are recorded at the lowest sampling rate by subsampling. Lu et al. [14] proposed a multi-rate dynamic modeling method for quality prediction at a faster rate based on the multiway partial least squares (PLS), in which the original dual sampled data is transformed into a three-dimensional matrix. Marjanovic et al. [15] presented a real-time monitoring scheme of an industrial batch process using the down-sampling methods. However, there are still some limitations for these methods. The prediction accuracy of the up-sampling methods relies on the regression model and the down-sampling methods result in loss of significant data information and distortion of process dynamics.

Therefore, it is more appropriate to directly use the multi-sampling-rate measurements without down-sampling or up-sampling. Data fusion techniques have provided an effective way for directly using multi-sample rate data [16]. Under the conditions of multi-rate sampling, the data fusion technology based on Kalman filter is introduced into soft sensor maintenance, and the fusion of soft sensor model estimation and process measurement is realized [17]. Huang et al. [18] have used deep learning (DL) methods to fuse signals with different sampling rates and proposed a multi-rate sampling data fusion method for fault diagnosis. As an alternative, the semi-supervised methods have been used for quality prediction in the dual-rate process [19], [20]. Zhu et al. [21] proposed a semi-supervised learning approach based on quantum statistic for industrial soft sensor development. Shao et al. [22] developed a semi-supervised Dirichlet mixture of Gaussians models, in which a fully Bayesian model structure is designed to implement semi-supervised tasks. Jin et al. [23] presented a semi-supervised soft sensor using evolutionary optimization-based pseudo labeling. Lima et al. [24] developed an industrial semi-supervised dynamic soft-sensor modeling approach based on deep relevant representation learning. Unfortunately, these semi-supervised methods only consider sampling rate inconsistencies between quality variables and process variables, and none of them can handle process data with three or more sampling rates. Recently, some of the semi-supervised probabilistic models have been extended to the multi-rate form. The multi-rate probabilistic principal component analysis (PPCA) and multi-rate factor analysis (MRFA) models have been developed for fault detection in multi-rate processes [25], [26]. Moreover, Zhou et al. proposed a multi-rate principal component regression model (MRPCR) for soft sensor applications in chemical process [27].

Besides, in order to improve the prediction performance of the model for multi-rate process, it is necessary to consider the process dynamics. However, the traditional dynamic process modeling methods focus on uniformly sampled data sets [28], [29]. To address this issue, the state of each measurement type is estimated using two Kalman filters, and the estimates are fused considering the correlation between them in the next step [30]. Aiming at the dual-rate characteristic of the system, Cao et al. [31] used the separation of multi-dynamic and static characteristics to predict the quality variable. But these models based on dual-rate system cannot be extended to three or more sampling rate systems. Furthermore, as a dynamic extension of the MRFA model, Cong et al. [32] proposed a multi-rate linear Gaussian state space model (MLGSS) for dynamic process monitoring. Although MLGSS can effectively extract dynamic latent variables in the multi-rate process, the model does not consider any output information. In the above literature, multi-rate dynamic modeling has found applications in fault detection and diagnosis. Its application to soft sensors has only gained recent attention with few industrial applications. Moreover, there are still some aspects which need to be improved. Firstly, we intend to extract the latent features of multi-rate processes that change slowly and reflect the internal trend of the process, because noise is usually included in the rapidly changing features. Secondly, in order to make full use of process information for quality prediction, the output information of the process is taken into account during the latent variable extraction. Finally, a more common model structure, model training procedure and the corresponding quality prediction strategy should be derived for any multi-rate dynamic processes.

To address these issues, the traditional probabilistic slow feature analysis is utilized and extended to its multi-rate form. And a multi-rate probabilistic slow feature regression (MR-PSFR) model is proposed for dynamic feature learning and industrial soft sensor development in this paper. Slow feature analysis (SFA) and probabilistic SFA (PSFA) are effective tools for dynamic modeling of industrial processes. By extracting latent features that vary slowly in time (i.e., slow features), they are able to capture all the dynamic information contained in the observations during modeling. Thus, a dimensionality reduction model with less noise can be obtained [33], [34], [35]. In the proposed MR-PSFR model, the whole input and output observations with different sampling rates are used to extract slow features (SFs), where both the internal correlations between the individual sampling rate and the internal correlations between different sampling rates are all considered. Moreover, the SFs extracted from MR-PSFR can separate slowly and fast changing latent features, and they can have a better interpretation of the outputs. Next, the Expectation-Maximization (EM) algorithm is modified to derive the model parameters of MR-PSFR. Finally, based on the established model, the quality prediction strategy for multi-rate processes is constructed.

The rest of the paper is organized as follows. Section 2 gives a brief introduction of SFA and PSFA. Section 3 presents a detailed explanation of the MR-PSFR model, followed by the model parameter estimation using EM method and the corresponding quality prediction strategy. Section 4 shows two case studies on a numerical example and an industrial process application. Finally, conclusion is made in Section 5.

II. RELATED WORK

A. SLOW FEATURE ANALYSIS

SFA is an unsupervised learning method proposed by Wiskott and Sejnowski [36]. The core idea of SFA is to extract some of the most slowly changing components from the time series as essential features. It aims to find a set of nonlinear functions $\{g_j(\cdot), 1 \le j \le q\}$ to map a d-dimensional input signal $\mathbf{x}(t)$ to a q-dimensional feature space. The outputs of these functions are called SFs, which are denoted as $s_j(t) := g_j(\mathbf{x}(t))$ $(1 \le j \le q)$. The SFA algorithm is to solve the following optimization problem:

$$\min_{g_j(\cdot)} \Delta(\cdot) := \min_{g_j(\cdot)} \left\langle \dot{s}_j^2(t) \right\rangle_t \tag{1}$$

subject to

$$\left\langle s_j(t) \right\rangle_t = 0, \quad \text{(zero mean)} \tag{2}$$

$$\left\langle s_j^2(t) \right\rangle_t = 1, \quad \text{(unit variance)} \tag{3}$$

$$\forall i \ne j, \left\langle s_i(t) s_j(t) \right\rangle_t = 0, \quad \text{(decorrelation and order)} \tag{4}$$

where $\Delta(\cdot)$ represents an indicator of changing rate of features, $\dot{s}_j(t) = s_j(t) - s_j(t-1)$ means the first-order time difference of $s(t)$, and $\langle\cdot\rangle_t$ denotes the expectation over time. Constraints (2) and (3) ensure the normalization of the output signal and avoid the appearance of the constant value solution, whereas constraint (4) ensures that each component of the output signal is uncorrelated and avoids redundant signals.

When the mapping functions $\{g_j(\cdot), 1 \le j \le q\}$ are linear, the SFs can be derived in linear form:

$$\mathbf{s}(t) = \mathbf{W}^{\mathrm{T}} \mathbf{x}(t) \tag{5}$$

where $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots \mathbf{w}_q] \in \mathbb{R}^{d \times q}$ is the mapping matrix. When the number of SFs is the same as that of inputs, i.e. $d = m$, the above optimization solution problem is equivalent to the following generalized eigenvalue decomposition problem:

$$\left\langle \dot{\mathbf{x}} \dot{\mathbf{x}}^T \right\rangle_t \mathbf{W} = \left\langle \mathbf{x} \mathbf{x}^T \right\rangle_t \mathbf{W} \Omega \tag{6}$$

where Ω is a diagonal matrix that contains the generalized eigenvalues $\{\omega_j\}$ on its diagonal, which are exactly the optimal values of objectives in Eq.(1), that is, $\Delta(\cdot) = \omega_j$. $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots \mathbf{w}_q]$ is the corresponding generalized eigenvector matrix [37].

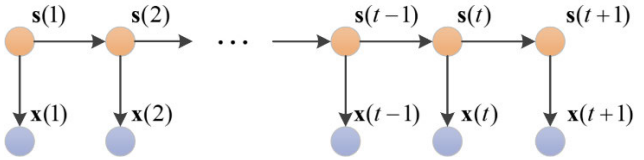


FIGURE 1. The model structure of PSFA.

B. PROBABILISTIC SFA

SFA has been extended in a probabilistic framework to PSFA by Turner and Sahani [38]. The model structure of PSFA is depicted in Figure 1 and the model formula is given as follows:

$$\begin{cases} \mathbf{s}(t) = \mathbf{F}\mathbf{s}(t-1) + \mathbf{e}_{\mathbf{s}}(t) \\ \mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{e}_{\mathbf{x}}(t) \end{cases} \tag{7}$$

where the state transition matrix $\mathbf{F}$ is defined as $\mathbf{F} = \mathrm{diag}\{\lambda_1, \cdots, \lambda_q\}$, $\mathbf{H} \in R^{d \times q}$ is the emission matrix, $\mathbf{e}_{\mathbf{s}}(t)$ is the state noise and follows a Gaussian distribution $\mathbf{e}_s(t) \sim N(\mathbf{0}, \mathbf{\Lambda})$, where the covariance matrix $\mathbf{\Lambda} = \mathrm{diag}\{1 - \lambda_1^2, \cdots, 1 - \lambda_q^2\}$, $\mathbf{e}_{\mathbf{x}}(t)$ is the observation noise and follows a Gaussian distribution $\mathbf{e}_{\mathbf{x}}(t) \sim N(\mathbf{0}, \mathbf{\Sigma})$, where the covariance matrix $\mathbf{\Sigma} = \mathrm{diag}\{\sigma_1^2, \cdots, \sigma_d^2\}$ and σ_j^2 is the variance of the jth dimension SF.

The independence assumption of SFs reflects the decorrelation nature of constraint (4). It can also be verified that the SFs derived from PSFA still satisfy the constraints (2) and (3), which are

$$E[s_j(t)] = 0,\ Var\{s_j(t)\} = 1,\ 1 \le j \le q \tag{8}$$

The transition parameter λ_j, which satisfies $0 \le \lambda_j < 1$, controls the correlation level between adjacent data points $s_j(t)$ and $s_j(t-1)$. In fact, the indicator $\Delta(\cdot)$ is calculated as $\Delta(s_j) = 2(1 - \lambda_j)$, which means that the larger λ_j is, the stronger the correlation between $s_j(t)$ and $s_j(t-1)$ is. $s_j(t)$ tends to have slower variation with a smaller $\Delta(\cdot)$, and vice versa.

As can be seen from Figure 1, the temporally correlated latent variables $\mathbf{s}(t)$ are derived from input variables $\mathbf{x}(t)$, which can only capture the dynamics in $\mathbf{x}(t)$ [33], [39]. It is noticed that the output observations could also contain dynamics which are beneficial to the prediction of future outputs. Moreover, the PSFA is trained using the uniformly sampling data. When faced with the problem of multiple sampling rates, quality prediction based on the PSFA model is not satisfactory. Based on the above considerations, the traditional PSFA is extended to its multi-sampling-rate form with consideration of output information in the next section.

III. MR-PSFR BASED QUALITY PREDICTION METHOD

A. MULTI-RATE PROBABILISTIC SLOW FEATURE REGRESSION MODEL

Suppose a multi-rate process contains M kinds of sampling rates for the process variables $\mathbf{X}$ and N kinds of sampling rates for the quality variables $\mathbf{Y}$. Given a multi-rate data set $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_t \in R^t, \mathbf{y}_t \in R^t\}_{t=1}^T$, the multi-rate probabilistic

slow feature regression model (MR-PSFR) is given as

$$\begin{cases} \mathbf{s}_t = \mathbf{F}\mathbf{s}_{t-1} + \mathbf{e}_\mathbf{s} \\ \mathbf{x}_t^{(m)} = \mathbf{H}^{(m)}\mathbf{s}_t + \mathbf{e}_\mathbf{x}^{(m)},\ m = 1, 2, \cdots, M \\ \mathbf{y}_t^{(n)} = \mathbf{U}^{(n)}\mathbf{s}_t + \mathbf{e}_\mathbf{y}^{(n)},\ n = 1, 2, \cdots, N \end{cases} \tag{9}$$

in which $\mathbf{x}_t^{(m)}$ and $\mathbf{y}_t^{(n)}$ denote observations at sample time t from different sampling rates. The observations of process variables at sampling time are denoted as $\mathbf{x}_t = \left[\mathbf{x}\left(o_t^1\right)^T\ \mathbf{x}\left(o_t^2\right)^T\ \cdots\ \mathbf{x}\left(o_t^{k_t}\right)^T\right]^T$, in which o_t^1 to $o_t^{k_t}$ represent different sampling rates and it is readily to obtain that $1 \le o_t^1 < o_t^2 < \cdots < o_t^{k_t} \le M$. Similarly, the measurements of quality variables are denoted as $\mathbf{y}_t = \left[\mathbf{y}\left(b_t^1\right)^T\ \mathbf{y}\left(b_t^2\right)^T\ \cdots\ \mathbf{y}\left(b_t^{J_t}\right)^T\right]^T$, in which b_t^1 to $b_t^{J_t}$ represents different sampling rates of $\mathbf{y}_t$, and $1 \le b_t^1 < b_t^2 < \cdots < b_t^{J_t} \le N$. The key factor of the MR-PSFR model is the slow feature $\mathbf{s}_t \in R^q$, which follows Gaussian distribution. The slow feature $\mathbf{s}_t$ is determined and shared by all the multi-rate measurements. $\mathbf{F} = \text{diag}\{\lambda_1, \cdots, \lambda_q\}$ is the state transition matrix, $\mathbf{H}^{(m)} \in R^{d(m)\times q}$ $(m = 1, 2, \cdots, M)$ are the emission matrix of the process variables under each sampling rate, and $\mathbf{U}^{(n)} \in R^{d(n)\times q}$ $(n = 1, 2, \cdots, N)$ are the emission matrix of the quality variables under each sampling rate, where $d(m)$ and $d(n)$ represent the variable dimensions of process variables and quality variables at different sampling rates, respectively. The state noise $\mathbf{e}_\mathbf{s}$ follows a Gaussian distribution $\mathbf{e}_\mathbf{s} \sim N(\mathbf{0}, \mathbf{\Lambda})$, where the covariance matrix $\mathbf{\Lambda} = \text{diag}\{1-\lambda_1^2, \cdots, 1-\lambda_q^2\}$. $\mathbf{e}_\mathbf{x}^{(m)}$ and $\mathbf{e}_\mathbf{y}^{(n)}$ are Gaussian noises of the process variables and quality variables, which are $\mathbf{e}_\mathbf{x}^{(m)} \sim N(\mathbf{0}, \Sigma_\mathbf{x}^{(m)})$, $m = 1, 2, \cdots, M$ and $\mathbf{e}_\mathbf{y}^{(n)} \sim N(\mathbf{0}, \Gamma_\mathbf{y}^{(n)})$, $n = 1, 2, \cdots, N$, where $\Sigma_\mathbf{x}^{(m)}$ and $\Gamma_\mathbf{y}^{(n)}$ are the noise variances.

The model structure diagram of MR-PSFR is depicted in Figure 2, where $M = 2$ and $N = 2$ are specified as an example. The whole observations with different sampling rates are used to extract SFs, where both the internal correlations between the individual sampling rate and the internal correlations between different sampling rates are all considered. Moreover, the extracted SFs take into account the output information of the process and can have a better interpretation of the output. The unknown model parameter set is $\Theta = \left\{\lambda_j, 1 \le j \le q,\ \mathbf{H}^{(m)}, \Sigma_\mathbf{x}^{(m)}, \mathbf{U}^{(n)}, \Gamma_\mathbf{y}^{(n)}\right\}$, $(m = 1, 2, \cdots, M;\ n = 1, 2, \cdots, N)$. In the next section, the EM algorithm is modified to derive the model parameters of MR-PSFR.

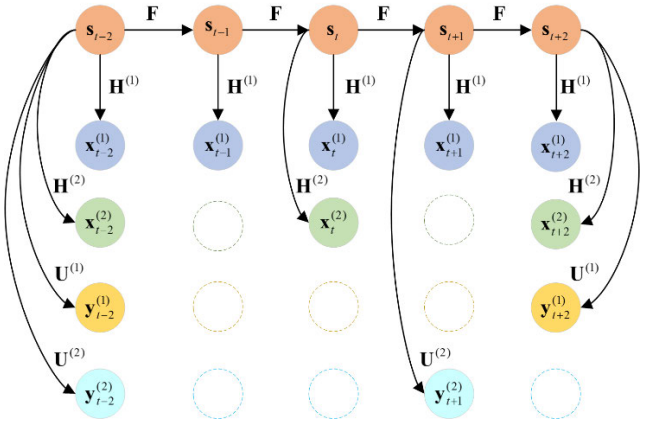


FIGURE 2. MR-PSFR model structure diagram.

B. MODEL PARAMETER ESTIMATION USING EM

The model parameters $\Theta = \left\{\lambda_j, 1 \le j \le q,\ \mathbf{H}^{(m)}, \Sigma_\mathbf{x}^{(m)}, \mathbf{U}^{(n)}, \Gamma_\mathbf{y}^{(n)}\right\}$, $(m = 1, 2, \cdots, M;\ n = 1, 2, \cdots, N)$ in MR-PSFR are derived using the EM algorithm. Given observations $D_o = \{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_1, \cdots, \mathbf{x}_T, \mathbf{y}_1, \cdots, \mathbf{y}_T\}$ and latent variables $D_{hid} = \{\mathbf{s}_1, \cdots, \mathbf{s}_T\}$, where $\mathbf{x}_t = \left[\mathbf{x}\left(o_t^1\right)^T\ \mathbf{x}\left(o_t^2\right)^T\ \cdots\ \mathbf{x}\left(o_t^{k_t}\right)^T\right]^T$ and $\mathbf{y}_t = \left[\mathbf{y}\left(b_t^1\right)^T\ \mathbf{y}\left(b_t^2\right)^T\ \cdots\ \mathbf{y}\left(b_t^{J_t}\right)^T\right]^T$, the global log-likelihood of the complete data is derived as (10), shown at the bottom of the page.

Assuming that the prior distribution of SFs is standard Gaussian distribution: $p(\mathbf{s}_1) = N(\mathbf{0}, \mathbf{I}_q)$, the first term in Eq. (10) is derived as

$$\ln p\left(\mathbf{s}_1\right) = \text{const} - \frac{1}{2}\mathbf{s}_1^T\mathbf{s}_1 \tag{11}$$

Then, the second term in Eq. (10) is calculated as

$$\begin{aligned} &\ln p\left(\mathbf{X}, \mathbf{Y}, \mathbf{s}|\Theta\right) \\ &= \ln \begin{cases} p\left(\mathbf{s}_1\right) p\left(\mathbf{x}_1^{(o_1^1)}, \cdots, \mathbf{x}_1^{(o_1^{k_1})}|\mathbf{s}_1\right) p\left(\mathbf{y}_1^{(b_1^1)}, \cdots, \mathbf{y}_1^{(b_1^{J_1})}|\mathbf{s}_1\right) \\ \prod_{t=2}^T p\left(\mathbf{s}_t|\mathbf{s}_{t-1}\right) p\left(\mathbf{x}_t^{(o_t^1)}, \cdots, \mathbf{x}_t^{(o_t^{k_t})}|\mathbf{s}_t\right) p\left(\mathbf{y}_t^{(b_t^1)}, \cdots, \mathbf{y}_t^{(b_t^{J_t})}|\mathbf{s}_t\right) \end{cases} \\ &= \ln p\left(\mathbf{s}_1\right) + \sum_{t=2}^T \ln p\left(\mathbf{s}_t|\mathbf{s}_{t-1}\right) + \sum_{t=1}^T \ln p\left(\mathbf{x}_t^{(o_t^1)}, \cdots, \mathbf{x}_t^{(o_t^{k_t})}|\mathbf{s}_t\right) \\ &\quad + \sum_{t=1}^T \ln p\left(\mathbf{y}_t^{(b_t^1)}, \cdots, \mathbf{y}_t^{(b_t^{J_t})}|\mathbf{s}_t\right) \end{aligned} \tag{10}$$

$$\sum_{t=2}^{T} \ln p\left(\mathbf{s}_t|\mathbf{s}_{t-1}\right) = \text{const} - \frac{T-1}{2}\sum_{j=1}^{q}\log\left(1-\lambda_j^2\right) - \frac{1}{2}\sum_{t=2}^{T}\sum_{j=1}^{q}\frac{1}{1-\lambda_j^2}\left(s_j(t) - \lambda_j s_j(t-1)\right)^2 \tag{12}$$

For the process variables, it can be proved that the covariance of the conditional distribution of the input observed samples with respect to the latent vector can be expressed as $\Sigma^{(o_t)} = \text{diag}\left(\Sigma_{\mathbf{x}}^{(o_t^1)};\ \Sigma_{\mathbf{x}}^{(o_t^2)};\ \cdots;\ \Sigma_{\mathbf{x}}^{(o_t^{k_t})}\right)$. Similarly, for the quality variables, the covariance of the conditional distribution of the output observed samples with respect to the latent vector is $\Gamma^{(b_t)} = \text{diag}\left(\Gamma_{\mathbf{y}}^{(b_t^1)};\ \Gamma_{\mathbf{y}}^{(b_t^2)};\ \cdots;\ \Gamma_{\mathbf{y}}^{(b_t^{J_t})}\right)$. Then, the third term and fourth term in Eq. (10) are derived as

$$\begin{aligned}&\sum_{t=1}^{T}\ln p\left(\mathbf{x}_t^{(o_t^1)},\cdots,\mathbf{x}_t^{(o_t^{k_t})}|\mathbf{s}_t\right)\\ &= \text{const} - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\Sigma^{(o_t)}\right| - \frac{1}{2}\sum_{t=1}^{T}\left[\left(\mathbf{x}_t - \mathbf{H}^{(o_t)}\mathbf{s}_t\right)^T\left(\Sigma^{(o_t)}\right)^{-1}\left(\mathbf{x}_t - \mathbf{H}^{(o_t)}\mathbf{s}_t\right)\right]\end{aligned} \tag{13}$$

$$\begin{aligned}&\sum_{t=1}^{T}\ln p\left(\mathbf{y}_t^{(b_t^1)},\cdots,\mathbf{y}_t^{(b_t^{J_t})}|\mathbf{s}_t\right)\\ &= \text{const} - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\Gamma^{(b_t)}\right| - \frac{1}{2}\sum_{t=1}^{T}\left[\left(\mathbf{y}_t - \mathbf{U}^{(b_t)}\mathbf{s}_t\right)^T\left(\Gamma^{(b_t)}\right)^{-1}\left(\mathbf{y}_t - \mathbf{U}^{(b_t)}\mathbf{s}_t\right)\right]\end{aligned} \tag{14}$$

in which the emission matrix of the process variables and the quality variables are defined as $\mathbf{H}^{(o_t)} = [\mathbf{H}^{(o_t^1)};\ \mathbf{H}^{(o_t^2)};\ \cdots;\ \mathbf{H}^{(o_t^{k_t})}]$ and $\mathbf{U}^{(b_t)} = [\mathbf{U}^{(b_t^1)};\ \mathbf{U}^{(b_t^2)};\ \cdots;\ \mathbf{U}^{(b_t^{J_t})}]$, respectively.

Substituting Eqs. (11)-(14) into Eq. (10), the Q-function can be formally derived by considering the conditional expectation of (10) as (15), shown at the bottom of the page, where $\boldsymbol{\Theta}^{\text{old}}$ represents the model parameters obtained in the previous iteration.

1) M-STEP

In the M-step, the new parameters $\boldsymbol{\Theta}^{\text{new}}$ can be calculated by maximizing the Q-function:

$$\Theta^{\text{new}} = \arg\max_{\theta} E_{\mathbf{s}|\mathbf{X},\mathbf{Y},\Theta^{\text{old}}}\left\{\ln p\left(\mathbf{X},\mathbf{Y},\mathbf{s}|\Theta\right)\right\} \tag{16}$$

Taking the derivative of the Q-function with respect to λ_j, we derive

$$\begin{aligned}\frac{\partial Q}{\partial\lambda_j} = \frac{(T-1)\lambda_j}{1-\lambda_j^2} - \frac{\lambda_j}{1-\lambda_j^2}\left\{\sum_{t=2}^{T}E\left[s_j^2(t-1)\right] - \sum_{t=2}^{T}E\left[s_j(t)s_j(t-1)\right]\right\}\\ - \frac{\lambda_j}{\left(1-\lambda_j^2\right)^2}\sum_{t=2}^{T}E\left[\left(s_j(t)-\lambda_j s_j(t-1)\right)^2\right] = 0\end{aligned} \tag{17}$$

Further, the updating equation (17) can be simplified as:

$$\begin{aligned}&(T-1)\lambda_j^3 - \sum_{t=2}^{T}E\left[s_j(t)s_j(t-1)\right]\cdot\lambda_j^2 + \left(\sum_{t=2}^{T}E\left[s_j^2(t)\right] + \sum_{t=2}^{T}E\left[s_j^2(t-1)\right] - T + 1\right)\lambda_j\\ &- \sum_{t=2}^{T}E\left[s_j(t)s_j(t-1)\right] = 0\end{aligned} \tag{18}$$

$$\begin{aligned}Q\left(\Theta,\Theta^{\text{old}}\right) &= E_{\mathbf{s}|\mathbf{X},\mathbf{Y},\Theta^{\text{old}}}\left\{\ln p\left(\mathbf{X},\mathbf{Y},\mathbf{s}|\Theta\right)\right\}\\ &= \text{const} - \frac{1}{2}\left\{\sum_{t=1}^{T}\ln\left|\Sigma^{(o_t)}\right| + \sum_{t=1}^{T}\ln\left|\Gamma^{(b_t)}\right|\right\}\\ &\quad - E\left\{\frac{T-1}{2}\sum_{j=1}^{q}\log\left(1-\lambda_j^2\right) + \frac{1}{2}\sum_{t=2}^{T}\sum_{j=1}^{q}\frac{1}{1-\lambda_j^2}\left(s_j(t)-\lambda_j s_j(t-1)\right)^2\right\}\\ &\quad - \frac{1}{2}\sum_{t=1}^{T}\left\{\begin{array}{l}\mathbf{x}_t^T(\Sigma^{(o_t)})^{-1}\mathbf{x}_t - 2(\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{x}_t E[\mathbf{s}_t]\\ +\text{tr}\left((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{H}^{(o_t)}E[\mathbf{s}_t\mathbf{s}_t^T]\right)\end{array}\right.\\ &\quad - \frac{1}{2}\sum_{t=1}^{T}\left\{\begin{array}{l}\mathbf{y}_t^T(\Gamma^{(b_t)})^{-1}\mathbf{y}_t - 2(\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{y}_t E[\mathbf{s}_t]\\ +\text{tr}\left((\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{U}^{(b_t)}E[\mathbf{s}_t\mathbf{s}_t^T]\right)\end{array}\right.\end{aligned} \tag{15}$$

By solving Eq. (18) and constraining the roots in the range [0,1], the updated λ_j can be obtained.

Parameters $\mathbf{H}^{(m)}$, $\Sigma_{\mathbf{x}}^{(m)}$, $\mathbf{U}^{(n)}$, $\Gamma_{\mathbf{y}}^{(n)}$ are updated by taking derivative of the Q-function with respect to each parameter and equating them to zero, which is given as

$$\mathbf{H}^{(m)\text{new}} = \left(\sum_{(m)} \mathbf{x}_t^{(m)} E[\mathbf{s}_t]\right)\left(\sum_{(m)} E[\mathbf{s}_t\mathbf{s}_t^T]\right)^{-1} \tag{19}$$

$$\Sigma_{\mathbf{x}}^{(m)\text{new}} = \frac{1}{N^{(m)}}\sum_{(m)}\left(\begin{array}{l}\mathbf{x}_t^{(m)}(\mathbf{x}_t^{(m)})^T - 2\mathbf{H}^{(m)\text{new}}E[\mathbf{s}_t](\mathbf{x}_t^{(m)})^T \\ +\mathbf{H}^{(m)\text{new}}E[\mathbf{s}_t\mathbf{s}_t^T](\mathbf{H}^{(m)\text{new}})^T\end{array}\right) \tag{20}$$

$$\mathbf{U}^{(n)\text{new}} = \left(\sum_{(n)} \mathbf{y}_t^{(n)} E[\mathbf{s}_t]\right)\left(\sum_{(n)} E[\mathbf{s}_t\mathbf{s}_t^T]\right)^{-1} \tag{21}$$

$$\Gamma_{\mathbf{y}}^{(n)\text{new}} = \frac{1}{N^{(n)}}\sum_{(n)}\left(\begin{array}{l}\mathbf{y}_t^{(n)}(\mathbf{y}_t^{(n)})^T - 2\mathbf{U}^{(n)\text{new}}E[\mathbf{s}_t](\mathbf{y}_t^{(n)})^T \\ +\mathbf{U}^{(n)\text{new}}E[\mathbf{s}_t\mathbf{s}_t^T](\mathbf{U}^{(n)\text{new}})^T\end{array}\right) \tag{22}$$

where $m = 1, 2, \cdots, M$; $n = 1, 2, \cdots, N$. $N^{(m)}$ and $N^{(n)}$ represent the sample numbers for the process variables and the quality variables at sampling rates m and n, respectively.

2) E-STEP

The parameter update formulas in M-step require three expectation terms: $E[\mathbf{s}_t]$, $E[\mathbf{s}_t\mathbf{s}_t^T]$ and $E[\mathbf{s}_t\mathbf{s}_{t-1}^T]$. The forward filtering and backward smoothing algorithm are used to estimate the expectations of SFs. For the multi-rate data set, some modifications to the forward filtering algorithm are required. Given the posterior distribution of the SF at time $t-1$, $p(\mathbf{s}_{t-1}|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = N(\mathbf{g}_{t-1}, \mathbf{G}_{t-1})$, in which $\mathbf{g}_{t-1}$ is the mean value and $\mathbf{G}_{t-1}$ is the covariance matrix, we need to calculate the posterior distribution of the SF at time t, $p(\mathbf{s}_t|\mathbf{x}_t, \mathbf{y}_t, \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1})$. According to the model structure of MR-PSFR, the joint probability distribution of $\mathbf{s}_t$, $\mathbf{x}_t$ and $\mathbf{y}_t$ can be obtained:

$$p(\mathbf{s}_t, \mathbf{x}_t, \mathbf{y}_t|\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) = N\left(\begin{bmatrix}\boldsymbol{\mu}_s \\ \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y\end{bmatrix}, \begin{bmatrix}\Sigma_{ss} & \Sigma_{sx} & \Sigma_{sy} \\ \Sigma_{xs} & \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{ys} & \Sigma_{yx} & \Sigma_{yy}\end{bmatrix}\right) \tag{23}$$

where $\boldsymbol{\mu}_s = \mathbf{F}\mathbf{g}_{t-1}$, $\boldsymbol{\mu}_x = \mathbf{H}^{(o_t)}\mathbf{F}\mathbf{g}_{t-1}$, $\boldsymbol{\mu}_y = \mathbf{U}^{(b_t)}\mathbf{F}\mathbf{g}_{t-1}$,
$\Sigma_{ss} = \mathbf{F}\mathbf{G}_{t-1}\mathbf{F}^T + \Lambda$, $\Sigma_{xx} = \mathbf{H}^{(o_t)}\Sigma_{ss}(\mathbf{H}^{(o_t)})^T + \Sigma^{(o_t)}$, $\Sigma_{yy} = \mathbf{U}^{(b_t)}\Sigma_{ss}(\mathbf{U}^{(b_t)})^T + \Gamma^{(b_t)}$, $\Sigma_{xs} = \mathbf{H}^{(o_t)}\Sigma_{ss}$, $\Sigma_{ys} = \mathbf{U}^{(b_t)}\Sigma_{ss}$, $\Sigma_{xy} = \mathbf{H}^{(o_t)}\Sigma_{ss}(\mathbf{U}^{(b_t)})^T$.

Based on the Bayes rule, the expectation and variance of the posterior distribution $p(\mathbf{s}_t|\mathbf{x}_t, \mathbf{y}_t, \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1})$ are derived as

$$\begin{aligned}\mathbf{g}_t = \boldsymbol{\mu}_s &+ \Big((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{H}^{(o_t)} \\ &+ (\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{U}^{(b_t)} + \Sigma_{ss}^{-1}\Big)^{-1} \\ &\times \Big((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_x) \\ &+ (\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_y)\Big)\end{aligned} \tag{24}$$

$$\begin{aligned}\mathbf{G}_t = &\Big((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{H}^{(o_t)} \\ &+ (\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{U}^{(b_t)} + \Sigma_{ss}^{-1}\Big)^{-1}\end{aligned} \tag{25}$$

The initial distribution parameters of the SF $\mathbf{g}_1$ and $\mathbf{G}_1$ can be given as

$$\begin{aligned}\mathbf{g}_1 = &\Big((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{H}^{(o_t)} \\ &+ (\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{U}^{(b_t)} + \mathbf{I}\Big)^{-1} \\ &\times \Big((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{x}_1 + (\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{y}_1\Big)\end{aligned} \tag{26}$$

$$\begin{aligned}\mathbf{G}_1 = &\Big((\mathbf{H}^{(o_t)})^T(\Sigma^{(o_t)})^{-1}\mathbf{H}^{(o_t)} \\ &+ (\mathbf{U}^{(b_t)})^T(\Gamma^{(b_t)})^{-1}\mathbf{U}^{(b_t)} + \mathbf{I}\Big)^{-1}\end{aligned} \tag{27}$$

Subsequently, the backward smoothing is performed according to the following formulas:

$$\hat{\mathbf{g}}_t = \mathbf{G}_t\mathbf{F}^T\Sigma_{ss}^{-1}(\hat{\mathbf{g}}_{t+1} - \mathbf{F}\mathbf{g}_t) + \mathbf{g}_t \tag{28}$$

$$\hat{\mathbf{G}}_t = \mathbf{G}_t\mathbf{F}^T\Sigma_{ss}^{-1}(\hat{\mathbf{G}}_{t+1}\Sigma_{ss}^{-1} - \mathbf{I})\mathbf{F}\mathbf{G}_t + \mathbf{G}_t \tag{29}$$

Finally, the evaluation of the three expectation terms are obtained as

$$E[\mathbf{s}_t] = \hat{\mathbf{g}}_t \tag{30}$$

$$E[\mathbf{s}_t\mathbf{s}_t^T] = \hat{\mathbf{G}}_t + \hat{\mathbf{g}}_t\hat{\mathbf{g}}_t^T \tag{31}$$

$$E[\mathbf{s}_t\mathbf{s}_{t-1}^T] = \mathbf{G}_{t-1}\mathbf{F}^T\Sigma_{ss}^{-1}\hat{\mathbf{G}}_t + \hat{\mathbf{g}}_t\hat{\mathbf{g}}_{t-1}^T \tag{32}$$

By iteratively updating and recalculating the E-step and M-step until convergence, the optimal model parameter set can be obtained. The latent state SFs can be inferred as the mean value $\mathbf{g}_t$ of the conditional probability distribution $p(\mathbf{s}_t|\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = N(\mathbf{g}_t, \mathbf{G}_t)$.

C. MR-PSFR BASED QUALITY PREDICTION

Based on the MR-PSFR model, the quality prediction strategy for multi-rate processes is proposed. Suppose we have obtained an online data sample $\mathbf{x}_{\text{new}}$ at time t_{new} and it contains R different sample rates:

$$\mathbf{x}_{\text{new}} = \left[\left(\mathbf{x}_{\text{new}}^{(o_{\text{new}}^1)}\right)^T \left(\mathbf{x}_{\text{new}}^{(o_{\text{new}}^2)}\right)^T \cdots \left(\mathbf{x}_{\text{new}}^{(o_{\text{new}}^R)}\right)^T\right]^T \tag{33}$$

in which $1 \le o_{\text{new}}^1 < o_{\text{new}}^2 < \cdots < o_{\text{new}}^R \le M$.

The online emission matrix and the online covariance matrix are constructed as

$$\mathbf{H}^{(o_{\text{new}})} = [\mathbf{H}^{(o_{\text{new}}^1)}; \mathbf{H}^{(o_{\text{new}}^2)}; \cdots; \mathbf{H}^{(o_{\text{new}}^R)}] \tag{34}$$

$$\Sigma^{(o_{\text{new}})} = \text{diag}\left(\Sigma_{\mathbf{x}}^{(o_{\text{new}}^1)}; \Sigma_{\mathbf{x}}^{(o_{\text{new}}^2)}; \cdots; \Sigma_{\mathbf{x}}^{(o_{\text{new}}^R)}\right) \tag{35}$$

It is assumed that the SFs in the previous sampling time follows a Gaussian distribution, which can be denoted as $N(\mathbf{s}_{\text{prev}}|\mathbf{g}_{\text{prev}}, \mathbf{G}_{\text{prev}})$. Then, according to the forward filtering

algorithm, the prediction of the SFs at time t_{new} is obtained as

$$\begin{aligned}\hat{\mathbf{s}}_{\text{new}} &= \mathbf{F}\mathbf{g}_{\text{prev}} + \Sigma_{ss}(\mathbf{H}^{(o_{\text{new}})})^T \\ &\times \left(\mathbf{H}^{(o_{\text{new}})}\Sigma_{ss}(\mathbf{H}^{(o_{\text{new}})})^T + \Sigma^{(o_{\text{new}})}\right)^{-1} \\ &\times (\mathbf{x}_{\text{new}} - \mathbf{H}^{(o_{\text{new}})}\mathbf{F}\mathbf{g}_{\text{prev}}) \end{aligned} \tag{36}$$

where $\Sigma_{ss} = \mathbf{F}\mathbf{G}_{\text{prev}}\mathbf{F}^T + \Lambda$.

By substituting Eq. (36) into the output equation in Eq. (9), the prediction of quality variables at time t_{new} can be estimated as

$$\mathbf{y}_{\text{new}}^{(n)} = \mathbf{U}^{(n)}\hat{\mathbf{s}}_{\text{new}}, \quad n = 1, 2, \cdots, N \tag{37}$$

Thus, the quality variables at time t_{new} for different sampling rates is obtained. To quantitatively evaluate the performance of the soft sensor model, the root mean square error (RMSE) is used as an assessment index, which is defined as

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(y_i - \hat{y}_i)^2} \tag{38}$$

where K is the size of the testing dataset, y_i is the real value of the output variable and $\hat{y}_i$ is the predicted value of the output variable.

IV. CASE STUDY

In this section, the proposed MR-PSFR method is applied to a numerical example and a R2S anaerobic reactor unit in the wastewater treatment process.

A. NUMERICAL EXAMPLE

A multi-rate dynamic model is simulated, in which the hidden variable $\mathbf{h}(t)$ is two-dimensional. The observed process variables $\mathbf{x}(t)$ include three sampling rates, and there are four process variables at each sampling rate. The observed quality variables $\mathbf{y}(t)$ include two sampling rates and two quality variables at each sampling rate. The simulated model is set as follows:

$$\begin{cases} \mathbf{h}(t+1) = \mathbf{A}\mathbf{h}(t) + \mathbf{e}_{\mathbf{h}}(t+1) \\ \mathbf{x}(t+1) = \mathbf{H}\mathbf{h}(t+1) + \mathbf{e}_{\mathbf{x}}(t+1) \\ \mathbf{y}(t+1) = \mathbf{U}\mathbf{h}(t+1) + \mathbf{e}_{\mathbf{y}}(t+1) \end{cases} \tag{39}$$

in which $t = 1, 2, \cdots, T-1$ is the sampling time, $\mathbf{e}_{\mathbf{h}}(t)$, $\mathbf{e}_{\mathbf{x}}(t)$ and $\mathbf{e}_{\mathbf{y}}(t)$ are Gaussian white noise with zero means and variance 0.01. It is defined as $\mathbf{H} = [\mathbf{H}_1;\ \mathbf{H}_2;\ \mathbf{H}_3]$ and $\mathbf{U} = [\mathbf{U}_1; \mathbf{U}_2]$, which are given as follows:

$$\mathbf{A} = \begin{bmatrix} 0.8 & 0.5 \\ 0.4 & -0.6 \end{bmatrix}, \quad \mathbf{U}_1 = \begin{bmatrix} 0.87 & 0.52 \\ 0.49 & 0.40 \end{bmatrix},$$

$$\mathbf{U}_2 = \begin{bmatrix} -0.17 & 0.58 \\ -0.20 & 0.75 \end{bmatrix}$$

$$\mathbf{H}_1 = \begin{bmatrix} 0.49 & 0.71 \\ -0.01 & -1 \\ 0.3 & 0.72 \\ -0.12 & 1.1 \end{bmatrix}, \mathbf{H}_2 = \begin{bmatrix} 0.28 & -0.13 \\ 0.62 & 0.98 \\ 1.63 & -0.67 \\ 1.36 & -1.12 \end{bmatrix},$$

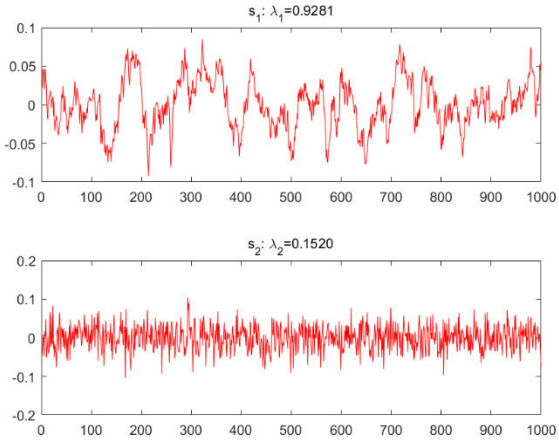


FIGURE 3. SFs extracted by MR-PSFR in numerical example.

TABLE 1. RMSE prediction errors of four models in numerical example.

Type	PPCR	PSFR	MRPCR	MR-PSFR
y_1	0.1936	0.1701	0.1467	0.1283
y_2	0.1441	0.1357	0.1288	0.1132
y_3	0.1444	0.1352	0.1229	0.1124
y_4	0.1450	0.1403	0.1321	0.1209

$$\mathbf{H}_3 = \begin{bmatrix} 1 & -1.13 \\ -0.5 & -0.64 \\ -0.62 & 0.08 \\ 1.56 & 0.62 \end{bmatrix} \tag{40}$$

Data series with 2000 samples are simulated from the abovementioned numerical setting. To carry out soft sensor modeling, the first 1000 samples are used as the training data while the rest 1000 samples are used for testing. Assuming that the sampling interval is T_s, process variables $x_1 \sim x_4$ are sampled every T_s, $x_5 \sim x_8$ are sampled every $2T_s$, and $x_9 \sim x_{12}$ are sampled every $5T_s$. For quality variables, $y_1 \sim y_2$ are sampled every $5T_s$, and $y_3 \sim y_4$ are sampled every $10T_s$. For performance comparison, the proposed MR-PSFR model is compared with the probabilistic principal components regression (PPCR) [40], probabilistic slow feature regression (PSFR) [33] and MRPCR [27] models. The MRPCR model is built using the same multi-rate dataset, while the PPCR and PSFR models are developed based on the down-sampling method. The latent variables' dimension of all the models is selected to be 2 with the numerical setting.

The two SFs extracted by MR-PSFR with corresponding λ_i are displayed in Figure 3, where λ_i reflects the slowness of the corresponding SF. The RMSE prediction errors for $y_1 \sim y_4$ using different models are presented in Table 1. It is obvious that the performance of MR-PSFR is superior to other approaches considered, which has the minimum RMSE for all outputs. In terms of y_1, the RMSE of the proposed model has percentage decreases of 33.73% compared to the PPCR model, percentage decreases of 24.57% compared to the PSFR model and percentage decreases of 12.54% compared to the MRPCR model. Figure 4 shows the detailed

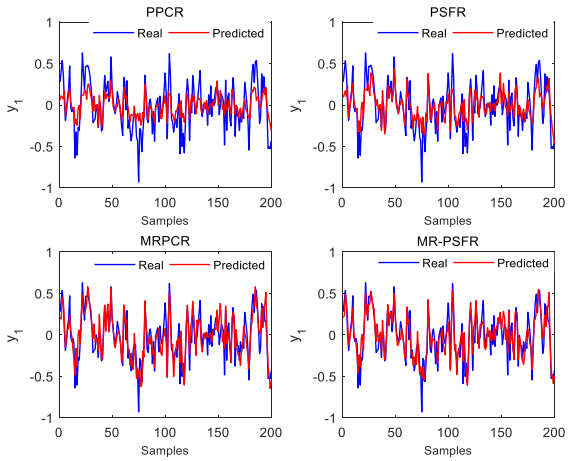


FIGURE 4. Prediction results for output y_1 in numerical example.

TABLE 2. Explanations of process variables in r2s reactor.

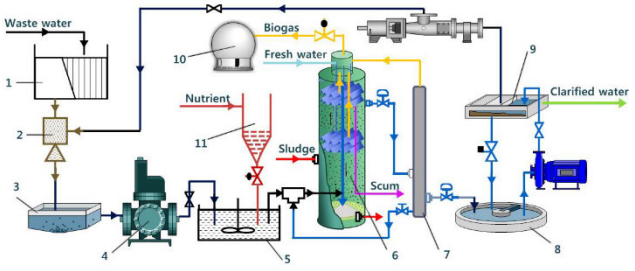
Tags	Explanation	Sampling frequency
U_1	Inlet flow rate in R2S-1	Online per hour
U_2	Inlet flow rate in R2S-2	Online per hour
U_3	Riser level of R2S	Online per hour
U_4	Water inflow of R2S-1	Online per 2 h
U_5	Water circulating flow of R2S-1	Online per 2 h
U_6	PH of inlet water in R2S-1	Online per 2 h
U_7	Temperature of inlet water in R2S-1	Online per 2 h
U_8	PH of outlet water in R2S-1	Online per 2 h
U_9	Temperature of outlet water in R2S-1	Online per 2 h
U_{10}	Water inflow of R2S-2	Online per 2 h
U_{11}	Water circulating flow of R2S-2	Online per 2 h
U_{12}	PH of inlet water in R2S-2	Online per 2 h
U_{13}	Temperature of inlet water in R2S-2	Online per 2 h
U_{14}	PH of outlet water in R2S-2	Online per 2 h
U_{15}	Temperature of outlet water in R2S-2	Online per 2 h
U_{16}	Inlet CODcr value in R2S-1	Off-line every 24 h
U_{17}	Inlet COD value in R2S-1	Off-line every 24 h
U_{18}	Inlet COD value in R2S-2	Off-line every 24 h
U_{19}	Inlet SS value in R2S-1	Off-line every 24 h
U_{20}	Inlet SS value in R2S-2	Off-line every 24 h
U_{21}	Inlet PH value in R2S-1	Off-line every 24 h
U_{22}	Inlet PH value in R2S-2	Off-line every 24 h

prediction results for output y_1 by the four models, where blue lines represent the real value and the red lines represent the predicted value. As can be seen from the figure, the prediction accuracy of MRPCR is better than that of PPCR and PSFR due to the use of multi-sampling rate datasets. Furthermore, the prediction performance has been greatly improved by the MR-PSFR model, especially for those samples 10th–17th, 100th–112th, and 178th–190th.

The experiments are performed on a personal computer with the configuration shown as follows. Operating system: Windows 10 (64-bit); CPU: Intel Core i7-10700 (2.90 GHz); RAM: 16.0 GB and MATLAB 2017b software. The computation time of the PPCR, PSFR, MRPCR and MR-PSFR models are 0.49s, 0.56s, 2.98s and 3.24s, respectively.

B. R2S ANAEROBIC REACTOR UNIT

The MR-PSFR model is applied to a real R2S anaerobic reactor unit in the papermaking wastewater treatment process. Figure 5 gives the detailed flowchart of the wastewater



1-Bar screener; 2-Blending pond; 3-Preliminary clarifier; 4-Pump; 5-Regulating container;6-Anaerobic reactor; 7-Cycling standpipe; 8-Anoxic pool; 9-The secondary clarifier; 10-Biogas storage tank; 11-Nutrient auto-count pipette

FIGURE 5. The flowchart of the wastewater treatment process.

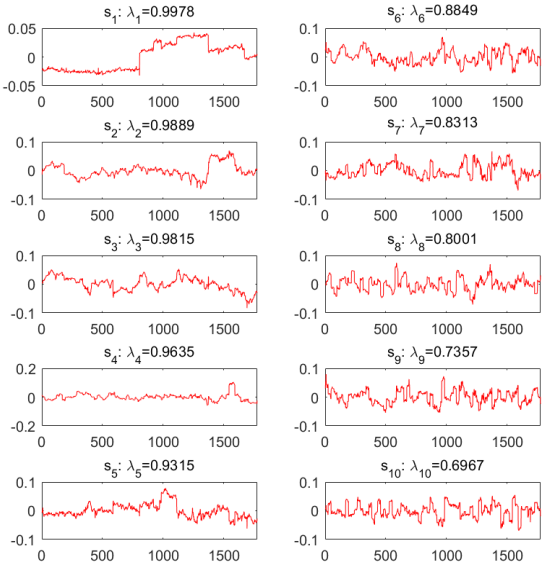


FIGURE 6. SFs extracted by MR-PSFR in wastewater treatment process.

treatment process. Papermaking wastewater treatment plant first carries out primary treatment for wastewater. The wastewater is passed successively through bar screener, blending pond and preliminary clarifier to remove large number of sediments, suspended solids and floating oil in the wastewater. Then it is fed into an anaerobic reactor for secondary treatment. The anaerobic and aerobic reaction of microorganisms is used to biodegrade wastewater. By adding suitable microbial products, the sludge yield and chemical Oxygen demand (COD) concentration can be significantly reduced. After secondary treatment, the wastewater goes through anoxic pool and secondary clarifier successively. Finally, it becomes clarified water to be recycled or directly discharged into the nearby water body.

Based on our engineering experience, 22 typical process variables are collected as the input data and five major quality indicators are chosen as the output data. Table 2 list the detailed process variable description. It can be seen that $U_1 \sim U_3$ are sampled online per hour, $U_4 \sim U_{15}$ are sampled online per 2 h and $U_{16} \sim U_{22}$ are off-line assayed every 24 h. Besides, the chosen output variables are chemical Oxygen demand (COD), volatile fatty acid (VFA), PH, suspended solids (SS) 1# and 2# in the anaerobic reactor outlet, all of

TABLE 3. RMSE prediction errors of four models in wastewater treatment process.

Type	PPCR	PSFR	MRPCR	MR-PSFR
COD	0.7225	0.5943	0.5716	0.4298
VFA	0.6891	0.5932	0.5420	0.4185
SS(1#)	0.4125	0.3850	0.3323	0.2841
SS(2#)	0.3601	0.3349	0.3118	0.2736

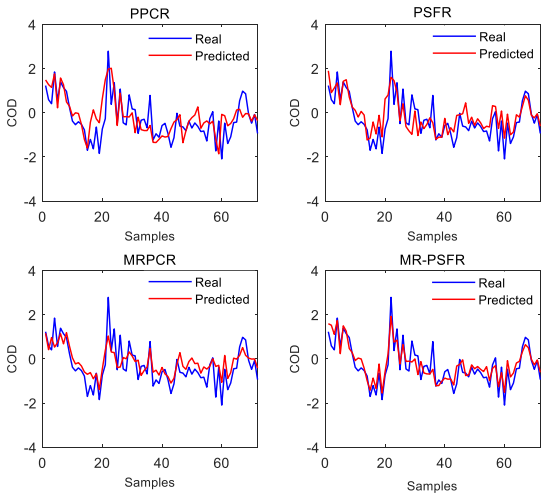


FIGURE 7. Prediction results for outlet COD in wastewater treatment process.

which are collected every 24 h by offline analyzing in the laboratory. Firstly, the data of 74 days under normal working conditions are collected as the training set and the MR-PSFR model is constructed. It can be found that 1776 samples of $U_1 \sim U_3$, 888 samples of $U_4 \sim U_{15}$, 74 samples of $U_{16} \sim U_{22}$ and 74 samples of the five quality indicators are collected. For comparison, the PPCR, PSFR and MRPCR models are also built using the same way as the numerical example. The latent variables of MR-PSFR and MRPCR are 10 and 8 respectively using the cross validation. And the latent variables of PPCR and PSFR are both 7. The ten slowest SFs extracted by MR-PSFR with corresponding λ_i are displayed in Figure 6, where λ_i reflects the slowness of the corresponding SF. It can be seen that the extracted SFs can separate slowly and fast changing latent features.

Subsequently, 72 days of test data from the same unit are collected to evaluate the predictive performance of the proposed model. The RMSE prediction errors for the five quality indicators using PPCR, PSFR, MRPCR and MR-PSFR are summarized in Table 3. Obviously, the proposed model performs better than the alternatives for all the five quality indicators. According to the prediction results of outlet COD, the RMSE of the proposed model has percentage decreases of 40.51% compared to the PPCR model, percentage decreases of 27.68% compared to the PSFR model and percentage decreases of 24.81% compared to the MRPCR model. The detailed prediction results for outlet COD using different models are illustrated in Figure 7. Note that the concentration of outlet COD is normalized. As can be seen from the figure, the prediction performance of MR-PSFR is the best among all models, and it can track the dynamic trend of the process. Especially for those samples 14th–19th, 21th–26th, and 55th–62th, the MR-PSFR has a far better prediction result than the alternatives. The reason is that the proposed method has utilized the sufficient process and quality variables instead of dropping some useful information, and the extracted SFs can capture all the dynamic information contained in observations during the modeling process. It can be inferred that the MR-PSFR model is an effective soft sensor for quality prediction in multi-rate dynamic processes.

V. CONCLUSION

In this paper, a multi-rate probabilistic slow feature regression model is proposed for dynamic feature learning and soft sensor development in industrial processes. In the MR-PSFR, both input and output observation datasets with different sampling rates are used to extract the slow features. Then, an efficient EM-based learning algorithm is developed for training the model and the quality prediction strategy for multi-rate processes is constructed based on MR-PSFR. Finally, two case studies are carried out on a numerical example and a real R2S anaerobic reactor unit in the wastewater treatment process. The simulation results show that the extracted slow features better represent the intrinsic characteristics of the processes and the proposed model has better prediction performance for multi-rate dynamic processes than other methods. In the future, more data-driven methods will be considered and applied to adapt to various irregular data sets in practice.

REFERENCES

[1] Z. Geng, Z. Chen, Q. Meng, and Y. Han, "Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1521–1529, Mar. 2022.

[2] B. Zhang, Y. Han, C. Li, and Z. Geng, "Novel gray orthogonal echo state network integrating the process mechanism for dynamic soft sensor development," *Ind. Eng. Chem. Res.*, vol. 60, no. 41, pp. 14955–14967, Oct. 2021.

[3] N. Yamada and H. Kaneko, "Adaptive soft sensor ensemble for selecting both process variables and dynamics for multiple process states," *Chemometrics Intell. Lab. Syst.*, vol. 219, Dec. 2021, Art. no. 104443.

[4] W. Xiong and W. Sun, "Dynamic soft sensor modeling for time sequence data based on convolution filters," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 6, pp. 2696–2703, Nov. 2021.

[5] Y. Liu, C. Yang, M. Zhang, Y. Dai, and Y. Yao, "Development of adversarial transfer learning soft sensor for multigrade processes," *Ind. Eng. Chem. Res.*, vol. 59, no. 37, pp. 16330–16345, Sep. 2020.

[6] S. Herceg, Ž. U. Andrijic, and N. Bolf, "Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models," *Chem. Eng. Res. Des.*, vol. 149, pp. 95–103, Sep. 2019.

[7] Y. Jiang, S. Yin, J. Dong, and O. Kaynak, "A review on soft sensors for monitoring, control, and optimization of industrial processes," *IEEE Sensors J.*, vol. 21, no. 11, pp. 12868–12881, Jun. 2020.

[8] S. Yin, Y. Li, B. Sun, Z. Feng, F. Yan, and Y. Ma, "Mixed kernel principal component weighted regression based on just-in-time learning for soft sensor modeling," *Meas. Sci. Technol.*, vol. 33, no. 1, Jan. 2022, Art. no. 015102.

[9] V. V. S. V. Krishna, N. Pappa, and S. P. J. Vasantharani, "Realization of deep learning based embedded soft sensor for bioprocess application," *Intell. Autom. Soft Comput.*, vol. 32, no. 2, pp. 781–794, 2022.
[10] J. Liu, J. Hou, and J. Chen, "Dual-layer feature extraction based soft sensor methods and applications to industrial polyethylene processes," *Comput. Chem. Eng.*, vol. 154, Nov. 2021, Art. no. 107469.
[11] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3721–3730, Jun. 2020.
[12] A. Khosbayar, J. Valluru, and B. Huang, "Multi-rate Gaussian Bayesian network soft sensor development with noisy input and missing data," *J. Process Control*, vol. 105, pp. 48–61, Sep. 2021.
[13] Z. Feng, Y. Li, B. Sun, C. Yang, and T. Huang, "A multimode mechanism-guided product quality estimation approach for multi-rate industrial processes," *Inf. Sci.*, vol. 596, pp. 489–500, Jun. 2022.
[14] N. Lu, Y. Yang, F. Gao, and F. Wang, "Multirate dynamic inferential modeling for multivariable processes," *Chem. Eng. Sci.*, vol. 59, no. 3, pp. 855–864, 2004.
[15] O. Marjanovic, B. Lennox, D. Sandoz, K. Smith, and M. Crofts, "Real-time monitoring of an industrial batch process," *Comput. Chem. Eng.*, vol. 30, nos. 10–12, pp. 1476–1481, Sep. 2006.
[16] M. A. K. Jaradat and R. Langari, "A hybrid intelligent system for fault detection and sensor fusion," *Appl. Soft Comput.*, vol. 9, no. 1, pp. 415–422, 2009.
[17] Y. Wu and X. Luo, "A novel calibration approach of soft sensor based on multirate data fusion technology," *J. Process Control*, vol. 20, no. 3, pp. 1252–1260, 2010.
[18] K. Huang, S. Wu, Y. Li, C. Yang, and W. Gui, "A multi-rate sampling data fusion method for fault diagnosis and its industrial applications," *J. Process Control*, vol. 104, pp. 54–61, Aug. 2021.
[19] J. M. M. de Lima and F. M. U. de Araujo, "Ensemble deep relevant learning framework for semi-supervised soft sensor modeling of industrial processes," *Neurocomputing*, vol. 462, pp. 154–168, Oct. 2021.
[20] L. Ren, Z. Meng, X. Wang, L. Zhang, and L. T. Yang, "A data-driven approach of product quality prediction for complex production systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6457–6465, Sep. 2021.
[21] J. Zhu, Z. Ge, and Z. Song, "Quantum statistic based semi-supervised learning approach for industrial soft sensor development," *Control Eng. Pract.*, vol. 74, pp. 144–152, May 2018.
[22] W. Shao, Z. Ge, and Z. Song, "Quality variable prediction for chemical processes based on semisupervised Dirichlet process mixture of Gaussians," *Chem. Eng. Sci.*, vol. 193, pp. 394–410, Jan. 2019.
[23] H. Jin, Z. Li, X. Chen, B. Qian, B. Yang, and J. Yang, "Evolutionary optimization based pseudo labeling for semi-supervised soft sensor development of industrial processes," *Chem. Eng. Sci.*, vol. 237, Jun. 2021, Art. no. 116560.
[24] J. M. M. de Lima and F. M. U. de Araújo, "Industrial semi-supervised dynamic soft-sensor modeling approach based on deep relevant representation learning," *Sensors*, vol. 21, no. 10, p. 3430, May 2021.
[25] L. Zhou, J. Chen, Z. Ge, and Z. Song, "Multiple fault detection using multi-rate probability principal component analysis models," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14752–14757, Jul. 2017.
[26] L. Zhou, Y. Wang, Z. Ge, and Z. Song, "Multi-rate factor analysis models for fault detection in multi-rate processes," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4076–4085, Jul. 2019.
[27] L. Zhou, Y. Wang, and Z. Ge, "Multi-rate principal component regression model for soft sensor application in industrial processes," *Sci. China Inf. Sci.*, vol. 63, no. 4, Apr. 2020, Art. no. 149205.
[28] Z. Ge and X. Chen, "Dynamic probabilistic latent variable model for process data modeling and regression application," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 1, pp. 323–331, Jan. 2019.
[29] X. Gao and Y. A. W. Shardt, "Dynamic system modelling and process monitoring based on long-term dependency slow feature analysis," *J. Process Control*, vol. 105, pp. 27–47, Sep. 2021.
[30] A. Fatehi and B. Huang, "Kalman filtering approach to multi-rate information fusion in the presence of irregular sampling rate and variable measurement delay," *J. Process Control*, vol. 53, pp. 15–25, May 2017.
[31] P. Cao, X. Luo, and X. Song, "Modeling and identification for soft sensor systems based on the separation of multi-dynamic and static characteristics," *Chin. J. Chem. Eng.*, vol. 26, no. 1, pp. 137–143, Jan. 2018.
[32] Y. Cong, L. Zhou, Z. Song, and Z. Ge, "Multirate dynamic process monitoring based on multirate linear Gaussian state-space model," *IEEE Trans. Automat. Sci. Eng.*, vol. 16, no. 4, pp. 1708–1719, Oct. 2019.
[33] C. Shang, B. Huang, F. Yang, and D. Huang, "Probabilistic slow feature analysis-based representation learning from massive process data for soft sensor modeling," *AIChE J.*, vol. 61, no. 12, pp. 4126–4139, 2015.
[34] L. Fan, H. Kodamana, and B. Huang, "Semi-supervised dynamic latent variable modeling: I/O probabilistic slow feature analysis approach," *AIChE J.*, vol. 65, no. 3, pp. 964–979, Mar. 2019.
[35] Y. Xu, M. Jia, and Z. Mao, "A novel auto-regressive dynamic slow feature analysis method for dynamic chemical process monitoring," *Chem. Eng. Sci.*, vol. 248, Feb. 2022, Art. no. 117236.
[36] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comput.*, vol. 14, no. 4, pp. 715–770, Apr. 2002.
[37] W. Zhong, C. Jiang, X. Peng, Z. Li, and F. Qian, "Online quality prediction of industrial terephthalic acid hydropurification process using modified regularized slow-feature analysis," *Ind. Eng. Chem. Res.*, vol. 57, no. 29, pp. 9604–9614, 2018.
[38] R. Turner and M. Sahani, "A maximum-likelihood interpretation for slow feature analysis," *Neural Comput.*, vol. 19, no. 4, pp. 1022–1038, Apr. 2007.
[39] Y. Qin, C. Zhao, and B. Huang, "A new soft-sensor algorithm with concurrent consideration of slowness and quality interpretation for dynamic chemical process," *Chem. Eng. Sci.*, vol. 199, pp. 28–39, May 2019.
[40] X. Yuan, Z. Ge, Z. Song, Y. Wang, C. Yang, and H. Zhang, "Soft sensor modeling of nonlinear industrial processes based on weighted probabilistic projection regression," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 4, pp. 837–845, Apr. 2017.

MIAO ZHANG received the B.Eng. and Ph.D. degrees from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2013 and 2018, respectively. She is currently a Lecturer with the School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou. Her research interests include industrial process modeling, soft sensor modeling, and machine learning.

ZHIWEI WEN received the B.Eng. degree from the Department of Industrial Automation, Wuhan Institute of Technology, Wuhan, China, in 1997. He is currently a Senior Engineer with Zhejiang Huashi Technology Company Ltd., Hangzhou, China, involved in intelligent equipment related works.

LE ZHOU (Member, IEEE) received the B.Eng. and Ph.D. degrees from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2010 and 2015, respectively.

He was a Visiting Scholar with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA, from December 2013 to June 2014, and a Visiting Scholar with the Department of Chemical Engineering, National Tsinghua University, Taiwan, from January 2019 to March 2019. He is currently an Associate Professor with the School of Automation and Electrical Engineering, Zhejiang University of Science and Technology. His research interests include industrial process modeling, monitoring and fault diagnosis, soft sensor modeling, and deep learning.