Kernel Generalization of Multi-Rate Probabilistic Principal Component Analysis for Fault Detection in Nonlinear Process

Donglei Zheng, Le Zhou, *Member, IEEE*, and Zhihuan Song

***Abstract*—In practical process industries, a variety of online and offline sensors and measuring instruments have been used for process control and monitoring purposes, which indicates that the measurements coming from different sources are collected at different sampling rates. To build a complete process monitoring strategy, all these multi-rate measurements should be considered for data-based modeling and monitoring. In this paper, a novel kernel multi-rate probabilistic principal component analysis (K-MPPCA) model is proposed to extract the nonlinear correlations among different sampling rates. In the proposed model, the model parameters are calibrated using the kernel trick and the expectation-maximum (EM) algorithm. Also, the corresponding fault detection methods based on the nonlinear features are developed. Finally, a simulated nonlinear case and an actual pre-decarburization unit in the ammonia synthesis process are tested to demonstrate the efficiency of the proposed method.**

***Index Terms*—Fault detection, kernel method, multi-rate process, probability principal component analysis (PPCA).**

I. Introduction

With the continuous developments of modern process industries and advances of data acquisition technologies, a large amount of industrial process data have been collected and stored by the distributed control system (DCS), which made data-driven process monitoring methods become a hot research topic in the field of process control in recent years [1]–[7]. Especially, with the rapid developments of machine learning and deep learning in recent decades, multivariate statistical process method (MSPM) technology can accurately express the operation state of processes at a low cost while also being easy to apply. Therefore, MSPM methods like principal component analysis (PCA), partial least squares (PLS) and their extensions have been widely applied in various types of process industries [8]–[12]. Besides, the process data are usually measured with different kinds of noises, which have an obvious impact on the data-driven model. Hence, the descriptions of the process noises should also be considered for process modeling purposes. At the same time, the process measurements are inherently stochastic variables instead of the deterministic values, which indicate that they are proper to be estimated in a probabilistic manner. Ge [13] has made a review of various types of probabilistic latent variable models in recent years, in which probabilistic PCA (PPCA), factor analysis (FA), and Gaussian mixture model (GMM) are widely used for process modeling and monitoring [14]–[16].

For most traditional MSPM methods, they all assumed that the sampled numbers of the training data for different variables are consistent. However, as the process industries have become more and more complex and production scales increase, the integrated automation system has been split into multiple levels and multiple sampling rates of the measurements are contained [17]–[19]. For example, the sampling rates of the process variables such as temperature, pressure, and flow are usually around a few seconds or minutes. Some important quality variables and other scheduling related data are tested at the laboratory with a much lower sampling rate, such as the product quality, exhaust gas concentration, and energy consumption index. They may be collected among a few hours or days. In general, various variables with different sampling rates from multiple levels can reflect the process status from different perspectives. Hence, it is desirable to utilize all the sampled measurements which are collected from multiple sampling rates so that different unfavorable conditions of processes like sensor faults, quality decline, or excessive energy consumption can be accurately detected and identified.

The traditional data-driven methods for multi-sampling rate process monitoring can be summarized as up-sampling methods, down-sampling methods, and semi-supervised methods. The main idea of the up-sampling method is to establish a regression model to predict the un-sampled data with a lower sampling rate. For comparison, the down-sampling method is to reduce all the variables to the lowest sampling rate to transfer a multiple sampling rate process to a single sampling rate process. After these two pretreatment technologies, the traditional MSPM method can be applied for

Manuscript received October 13, 2020; accepted November 20, 2020. This work was supported by Zhejiang Provincial Natural Science Foundation of China (LY19F030003), Key Research and Development Project of Zhejiang Province (2021C04030), the National Natural Science Foundation of China (62003306) and Educational Commission Research Program of Zhejiang Province (Y202044842). Recommended by Associate Editor M. Chadli. *(Corresponding author: Le Zhou.)*

Citation: D. L. Zheng, L. Zhou, and Z. H. Song, "Kernel generalization of multi-rate probabilistic principal component analysis for fault detection in nonlinear process," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 8, pp. 1465–1476, Aug. 2021.

D. L. Zheng and Z. H. Song are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: 21832086@zju.edu.cn; songzhihuan@zju.edu.cn).

L. Zhou is with School of Automation and Electrical Engineering, Zhejiang University of Science & Technology, Hangzhou 310023, China (e-mail: zhoule@zust.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2021.1004090

process modeling and monitoring. Lu *et al*. [20] transformed the original dual sampled data into a three-dimensional matrix by adjusting the sample structure and using a multiway PLS model for key variable prediction. Marjanovic *et al*. [21] extended the process variables with different sampling scales to the high-dimensional vectors according to the sampling interval of the quality variable. After that, the batch process monitoring scenario is established using the down-sampled data. Besides, Li *et al*. [22] used fast sampling and slow sampling samples through the lifting technology and resampling for batch process modeling. However, there are still several limitations to these methods. Up-sampling methods greatly depend on the accuracy of the regression model. Down-sampling methods break correlations of their original data structures. When the difference between sampling rates is too large, training data will be statistically insufficient and cannot derive an accurate enough model.

Another choice for multi-rate process modeling is to directly utilize the original dataset. The probabilistic model can combine the probability inference and the expectation maximization (EM) algorithm, which makes it possible for estimating the missing data and multi-rate process modeling [23]. Especially, the semi-supervised methods have been used for quality prediction and monitoring in the dual-rate process. Ge *et al*. [24] proposed a semi-supervised algorithm based on Bayesian canonical PPCA and a selection method of the dimension of latent variables. Zhou *et al*. [25] proposed a semi-supervised probabilistic latent variable regression (PLVR) model and applied it on a continuous process and batch process, in which the unbalanced fast sampling process variables and slow sampling quality indexes were both involved. Recently, the semi-supervised probabilistic model has been extended to the multi-rate form. The multi-rate PPCA and multi-rate factor analysis models are developed to incorporate the measurements with different sampling rates without down-sampling and up-sampling. Also, the common latent variables are extracted in the probabilistic framework to describe the auto-correlations among different sampling scales [26], [27].

Besides, the correlation relationships in the actual industrial processes are usually nonlinear, which indicates that linear models will be improper for these cases. To deal with the nonlinear relationship among process and quality variables, the kernel method has been widely used in many statistical learning methods such as support vector machines (SVM), kernel principal component analysis (KPCA), kernel probabilistic principal component analysis (KPPCA), kernel partial least squares (KPLS) [28]–[31], etc. Huang and Yan [32] proposed a quality-driven kernel principal component analysis (Q-KPCA) model for quality dependent nonlinear process monitoring by decomposing the kernel matrices into quality related and unrelated spaces. With the help of the kernel trick, the nonlinear correlations can be projected to be linear in the feature subspace. Besides, ensemble learning is also an idea for monitoring nonlinear processes. Li and Yan [33], [34] used the ensemble learning to solve nonlinear problems and improve monitoring performance. More recently, a deep learning technology has been rapidly developed with the enhancements of the computing hardware and the increase of the data storage capacity [35]. Using end-to-end learning, the parameters of feature extraction and pattern classification in the multi-hidden layer network can be coordinated and optimized. For process modeling and monitoring purposes, deep learning technologies have also been applied. Jiang *et al*. [36] proposed a unified robust self-monitoring framework, which adds the artificial data into the model input and introduces the robust training method into the self-supervised model. Deng *et al*. [37] proposed a deep learning method based on the nonlinear PCA model for industrial process monitoring and fault detection named deep PCA (DePCA). Wu and Zhao [38] performed the convolution operation on chemical industry data sets and proposed a deep convolutional neural network (DCNN) model for fault diagnosis.

In these models, the nonlinear activation functions are contained in each layer, which makes it possible to solve nonlinear problems and have good fitting abilities. However, most current deep learning models still require that the dimensions of input samples must be unified. It is hard to directly utilize the traditional deep learning models for the nonlinear multi-rate processes.

In past research mentioned above, those mainly focused on the model of single sampling rate, which led to the lack of process monitoring of multi-sampling rate data. Besides, the current multi-rate process modeling methods rarely considered the nonlinear constraint relationship. Hence, the kernel multi-rate probabilistic principal component analysis (K-MPPCA) model is proposed to extract the nonlinear correlations with different sampling rates. With the aid of the kernel trick, the common principal components are derived in the high-dimensional feature space, in which the correlations between the principal components and the projected measurements among different sampling rates are assumed to be linear. The model parameters can also be calibrated using the EM algorithm. Moreover, the fault detection scenarios with different sampling rates are further developed in the feature space and several residual spaces for each sampling rate. The corresponding fault detection performance is expected to be more accurate since the nonlinear correlations are extracted and monitored without destroying the original multi-rate structures.

The contribution of this work can be summarized as follows:

1) The constraint nonlinear relationship with different sampling rates are considered in the proposed K-MPPCA model. Besides, a nonlinear EM algorithm is proposed for model parameter estimation since the explicit nonlinear correlations between the original measurement and the latent variables cannot be obtained after the kernel method is introduced.

2) A fault detection scheme for a nonlinear multi-rate process is developed. In K-MPPCA, however, the squared prediction error (SPE) statistics are difficult to derive in high-dimensional feature space. Hence, it is further constructed with the help of the kernel trick in this paper.

The rest of this paper is organized as follows. The section "Revisit of MPPCA" briefly reviews the MPPCA model, which is linear and based on the whole multi-rate measurement. Then, the K-MPPCA model and its model

parameter estimation method for the nonlinear multi-rate process are introduced in detail in "K-MPPCA Model". "K-MPPCA based Fault Detection" implements the nonlinear process monitoring strategy using the multi-rate measurements. "Case Studies" demonstrates a numerical example and an actual industrial process to evaluate the performance of the proposed method. Finally, some conclusions are made.

II. Revisit of Multiple Probability Principal Component Analysis

As a multi-rate form of PPCA model, MPPCA has a similar structure to PPCA and can make full use of data with different sampling rates, which is given as follows

$$\begin{cases} \mathbf{x}_1 = \mathbf{\Phi}_1\mathbf{t}+\varepsilon_1 \\ \mathbf{x}_2 = \mathbf{\Phi}_2\mathbf{t}+\varepsilon_2 \\ \quad\vdots \\ \mathbf{x}_{S-1} = \mathbf{\Phi}_{S-1}\mathbf{t}+\varepsilon_{S-1} \\ \mathbf{x}_S = \mathbf{\Phi}_S\mathbf{t}+\varepsilon_S . \end{cases} \tag{1}$$

In this model, all the measurements with different sampling rates can be included in a unique model. $\{\mathbf{x}_1\ \mathbf{x}_2\ \cdots \mathbf{x}_{S-1}\ \mathbf{x}_S\}$ are the process variables, where sampling rate x_1 is the highest and sampling rate x_S is the lowest. In the actual industrial processes, some process variables are collected in just a few seconds and some key quality indexes need the offline test which takes a lot of time. Therefore, it can be assumed that the number of samples for process variables at different sampling rates K_1, K_2,…, K_{S-1} and K_S are given as $K_S < K_{S-1} < \cdots < K_2 < K_1$, and $\{\mathbf{\Phi}_1\ \mathbf{\Phi}_2\ \cdots \mathbf{\Phi}_{S-1}\ \mathbf{\Phi}_S\}$ are the loading matrices of the model under different sampling rates. Like most probability models, the latent variable $\mathbf{t}$ shared by $x_1, x_2, \ldots, x_{S-1}, x_S$ is assumed to follow a Gaussian distribution with zero mean and unit variance. $\varepsilon_1 \in \mathbb{R}^{M_1}$, $\varepsilon_2 \in \mathbb{R}^{M_2}$, …, $\varepsilon_{S-1} \in \mathbb{R}^{M_{S-1}}$ and $\varepsilon_S \in \mathbb{R}^{M_S}$ are used to describe their respective noises; they are defined to follow the isotropic Gaussian distribution as $\varepsilon_1 \sim N(0,\sigma_1^2\mathbf{I})$, $\varepsilon_2 \sim N(0,\sigma_2^2\mathbf{I})$, …, $\varepsilon_{S-1} \sim N(0,\sigma_{S-1}^2\mathbf{I})$ and $\varepsilon_S \sim N(0,\sigma_S^2\mathbf{I})$, respectively. The MPPCA model has brought about a possible solution to the multi-rate process modeling problem. However, it still only extracts the linear relationship among different sampling rates. Hence, it is desirable to derive its nonlinear form for the real multi-rate industrial processes.

III. Kernel Multiple Probability Principal Component Analysis Model

In the actual multi-rate process, process variables often exhibit a strong nonlinear relationship which leads to poor interpretation of the loading matrices (such as $\{\mathbf{\Phi}_1\ \mathbf{\Phi}_2\ \cdots\ \mathbf{\Phi}_{S-1}\ \mathbf{\Phi}_S\}$, as obtained in the MPPCA model) and has some limitations on the fault detection approach based on it. In this section, the linear MPPCA model is generalized to its nonlinear form using the kernel trick, which is named as the multi-rate probability kernel principal component analysis (K-MPPCA) model. Also, its model parameter estimation methods are derived.

A. Kernel Multiple Probability Principal Component Analysis Model (K-MPPCA)

The nonlinear relationship between the multi-rate measurements and their common latent variables are constructed in K-MPPCA. The original multi-rate measurements are $\{\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_{S-1}\ \mathbf{x}_S\}$, in which their dimensions are $\{M_1, M_2, \ldots, M_{S-1}, M_S\}$, respectively. In K-MPPCA, they are first projected into a high-dimensional feature space as $\mathbf{\Phi} : \mathbf{x}_s \in \mathbb{R}^{M_s} \mapsto \mathbf{\Phi}(\mathbf{x}_s) \in \mathcal{F}$, where the correlations between the latent variables and $\Phi(\mathbf{x}_s)$ are thought of as linear, and they are given as

$$\begin{cases} \Phi(\mathbf{x}_1) = \mathbf{P}_1\mathbf{t}+e_1 \\ \Phi(\mathbf{x}_2) = \mathbf{P}_2\mathbf{t}+e_2 \\ \quad\vdots \\ \Phi(\mathbf{x}_{S-1}) = \mathbf{P}_{S-1}\mathbf{t}+e_{S-1} \\ \Phi(\mathbf{x}_S) = \mathbf{P}_S\mathbf{t}+e_S \end{cases} \tag{2}$$

where the correlations between the latent variables and $\Phi(\mathbf{x}_s)$ are thought as linear and constructed using the loading matrices $\{\mathbf{P}_1\ \mathbf{P}_2 \ldots \mathbf{P}_{S-1}\ \mathbf{P}_S\}$. The common latent variable $\boldsymbol{t}$ is shared by all the measurements with different sampling rates. The noises $\{\mathbf{e}_1 \sim N(0,\sigma_1^2\mathbf{I}),\ \mathbf{e}_2 \sim N(0,\sigma_2^2\mathbf{I}), \ldots, \mathbf{e}_{S-1} \sim N(0,\sigma_{S-1}^2\mathbf{I}),\ \mathbf{e}_S \sim N(0,\sigma_S^2\mathbf{I})\}$ are assumed to be Gaussian and are used to describe the residual spaces of the model. The whole measurements which contain different classes of sampling rate are divided into different partitions. Therefore, all multi-rate measurements can be arranged and divided into S partitions as $\mathbf{V}_S = \{(\Phi(\mathbf{x}_{1k}), \Phi(\mathbf{x}_{2k}), \ldots, \Phi(\mathbf{x}_{Sk})) | k = 1, \ldots, K_S\}$, $\mathbf{V}_{S-1} = \{(\Phi(\mathbf{x}_{1k}), \Phi(\mathbf{x}_{2k}), \ldots, \Phi(\mathbf{x}_{S-1,k})) | k = K_S+1, \ldots, K_{S-1}\}$, …, $\mathbf{V}_2 = \{\Phi(\mathbf{x}_{1k}), \Phi(\mathbf{x}_{2k}) | k = K_3+1, \ldots, K_2\}$ and $\mathbf{V}_1 = \{\Phi(\mathbf{x}_{1k}) | k = K_2+1, \ldots, K_1\}$. According to the above arrangement, the whole data set can be written as a union as $\mathbf{V} = \mathbf{V}_S \cup \mathbf{V}_{S-1} \cup \cdots \cup \mathbf{V}_2 \cup \mathbf{V}_1$. The reason for dividing all data into S partitions is that the expected values of latent variables in different partitions are updated differently. At the same time, when updating model parameters $\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_S^2\}$, different σ^2 requires the expectation of the latent variables from different partitions. Therefore, dividing the data into S partitions can make the model more concise and the parameter updating procedures clearer. Although the data are divided into S partitions, the latent variables are jointly determined by all the multi-sampling rate data. Hence, the K-MPPCA model is still a unified model unlike those separated models.

B. Model Parameter Estimation Using Nonlinear EM

Like most probability models, the model parameters $\{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_S, \sigma_1^2, \sigma_2^2, \ldots, \sigma_S^2\}$ of K-MPPCA can be estimated by the EM algorithm, which calibrates the model parameters by iterating the expectation (E-step) and maximization (M-step) steps until the model parameters converge. In the E-step, the current model parameters are used to calculate the posterior distribution of the latent variables, which is given as

$$E(\hat{\mathbf{t}}_k \mid \mathbf{V}_i) = \mathbf{\Omega}_i^{-1} \sum_{s=1}^{i} \sigma_s^{-2}\mathbf{P}_s^T \Phi(\mathbf{x}_{s,k}),\ \ i = 1, 2, \ldots, S \tag{3}$$

$$\mathbf{\Omega}_i^{-1} = \sum_{s=1}^{i} \sigma_s^{-2}\mathbf{P}_s^T\mathbf{P}_s + \mathbf{I},\ \ i = 1, 2, \ldots, S \tag{4}$$

$$E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T \mid \mathbf{V}_i) = \mathbf{\Omega}_i^{-1} + E(\hat{\mathbf{t}}_k \mid \mathbf{V}_i) E^T(\hat{\mathbf{t}}_k \mid \mathbf{V}_i),\ \ i = 1, 2, \ldots, S . \tag{5}$$

In the M-step, the exponential likelihood function is maximized, which is given as

$$\begin{aligned} L &= L_S + L_{S-1} + \cdots + L_2 + L_1 \\ &= \sum_{k=1}^{K_S} \ln p(\Phi(\mathbf{x}_{1k}), \Phi(\mathbf{x}_{2k}), \ldots, \Phi(\mathbf{x}_{Sk})) \\ &\quad + \sum_{k=K_S+1}^{K_{S-1}} \ln p(\Phi(\mathbf{x}_{1k}), \Phi(\mathbf{x}_{2k}), \ldots, \Phi(\mathbf{x}_{S-1,k})) \\ &\quad + \cdots + \sum_{k=K_3+1}^{K_2} \ln p(\Phi(\mathbf{x}_{1k}), \Phi(\mathbf{x}_{2k})) + \sum_{k=K_2+1}^{K_1} \ln p(\Phi(\mathbf{x}_{1k})). \end{aligned} \tag{6}$$

The updated model parameters value can be obtained by taking the partial derivative of the likelihood function concerning the parameter and they are estimated as

$$\begin{aligned} \hat{\mathbf{P}}_i &= \left[\begin{array}{l} \displaystyle\sum_{k=1}^{K_S} \Phi(\mathbf{x}_{ik}) E^T\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_S\right) + \sum_{k=K_S+1}^{K_{S-1}} \Phi(\mathbf{x}_{ik}) E^T\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_{S-1}\right) \\ \quad + \cdots + \displaystyle\sum_{k=K_2+1}^{K_I} \Phi(\mathbf{x}_{ik}) E^T(\mathbf{t}_k \mid \mathbf{V}_i) \end{array} \right] \\ &\quad \times \left[\begin{array}{l} \displaystyle\sum_{k=1}^{K_S} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_S\right) + \sum_{k=K_S+1}^{K_{S-1}} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_{S-1}\right) \\ \quad + \cdots + \displaystyle\sum_{k=K_{i+1}+1}^{K_i} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_i\right) \end{array} \right]^{-1} \\ &\quad i = 1, 2, \ldots, S \end{aligned} \tag{7}$$

$$\begin{aligned} \hat{\sigma}_i^2 = \frac{1}{M_i K_i} \Bigg\{ & \sum_{k=1}^{K_1} \Phi(\mathbf{x}_{1k})^T \Phi(\mathbf{x}_{1k}) \\ & -2 \left(\begin{array}{l} \displaystyle\sum_{k=1}^{K_S} E^T\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_S\right) \hat{\mathbf{P}}_i^T \Phi(\mathbf{x}_{ik}) \\ + \displaystyle\sum_{k=K_S+1}^{K_{S-1}} E^T\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_{S-1}\right) \hat{\mathbf{P}}_1^T \Phi(\mathbf{x}_{ik}) \\ + \cdots + \displaystyle\sum_{k=K_{i+1}+1}^{K_i} E^T(\mathbf{t}_k \mid \mathbf{V}_i) \hat{\mathbf{P}}_i^T \Phi(\mathbf{x}_{ik}) \end{array} \right) \\ & + trace\left(\left(\begin{array}{l} \displaystyle\sum_{k=1}^{K_S} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_S\right) \\ + \displaystyle\sum_{k=K_S+1}^{K_{S-1}} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_{S-1}\right) \\ + \cdots + \displaystyle\sum_{k=K_i+1}^{K_i} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_i\right) \end{array} \right) \hat{\mathbf{P}}_i^T \hat{\mathbf{P}}_i \right) \Bigg\} \\ & i = 1, 2, \ldots, S. \end{aligned} \tag{8}$$

It can be readily obtained that the model parameters cannot be directly estimated using the above equations since the projected values $\{\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_{S-1}), \Phi(\mathbf{x}_S)\}$ are unknown. Therefore, it is necessary to introduce the kernel method by calculating the kernel matrix of different sampling variables. In such a manner, a nonlinear EM algorithm is developed.

Firstly, all the loading matrices can be decomposed into three parts, which are given as follows:

$$\hat{\mathbf{P}}_i = \Phi(\mathbf{X}_i)^T \mathbf{T}_i \mathbf{C}_i^{-1}, \ i = 1, 2, \ldots, S \tag{9}$$

where $\{\Phi(X_i), T_i, C_i\} (i = 1, 2 \ldots S)$ are respectively defined as

$$\begin{aligned} \Phi(\mathbf{X}_i) &= \left[\Phi(\mathbf{x}_{i,1}), \Phi(\mathbf{x}_{i,2}), \ldots, \Phi\left(\mathbf{x}_{iK_i-1}\right), \Phi\left(\mathbf{x}_{i,K_i}\right)\right]^T \\ &\quad i = 1, 2, \ldots, S \end{aligned} \tag{10}$$

$$\begin{aligned} \mathbf{T}_i &= \left[\mathbf{T}_{\mathbf{V}_S}, \mathbf{T}_{\mathbf{V}_{S-1}}, \ldots, \mathbf{T}_{\mathbf{V}_i}\right]^T \\ &= \left[\begin{array}{l} E\left(\hat{\mathbf{t}}_1 \mid \mathbf{V}_S\right), \ldots, E\left(\hat{\mathbf{t}}_{K_S} \mid \mathbf{V}_S\right) \\ E\left(\hat{\mathbf{t}}_{K_S+1} \mid \mathbf{V}_{S-1}\right), \ldots, E\left(\hat{\mathbf{t}}_{K_{S-1}} \mid \mathbf{V}_{S-1}\right) \\ \cdots \\ E\left(\hat{\mathbf{t}}_{K_{i+1}+1} \mid \mathbf{V}_2\right), \ldots, E\left(\hat{\mathbf{t}}_{K_i} \mid \mathbf{V}_2\right) \end{array} \right]^T \\ &\quad i = 1, 2, \ldots, S \end{aligned} \tag{11}$$

$$\begin{aligned} \mathbf{C}_i &= \sum_{S=i}^{S} \mathbf{C}_{\mathbf{V}_S} = \left[\begin{array}{l} \displaystyle\sum_{k=1}^{K_S} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_S\right) + \sum_{k=K_S+1}^{K_{S-1}} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_{S-1}\right) \\ \quad + \cdots + \displaystyle\sum_{k=K_{i+1}+1}^{K_i} E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_2\right) \end{array} \right] \\ &\quad i = 1, 2, \ldots, S. \end{aligned} \tag{12}$$

At the same time, the equivalent transformation can be obtained as

$$\begin{aligned} &\hat{\mathbf{P}}_i^T \hat{\mathbf{P}}_i = \mathbf{C}_i^{-1} \mathbf{T}_i^T \Phi(\mathbf{X}_i) \Phi(\mathbf{X}_i)^T \mathbf{T}_i \mathbf{C}_i^{-1} = \mathbf{C}_i^{-1} \mathbf{T}_i^T \mathbf{K}_i \mathbf{T}_i \mathbf{C}_i^{-1} \\ &\hat{\mathbf{P}}_i^T \Phi(\mathbf{x}_{ik}) = \mathbf{C}_i^{-1} \mathbf{T}_i^T \Phi(\mathbf{X}_i) \Phi(\mathbf{x}_{ik}) = \mathbf{C}_i^{-1} \mathbf{T}_i^T \mathbf{k}_{ik}, \ i = 1, 2, \ldots, S \end{aligned} \tag{13}$$

where we introduce the kernel trick $\mathbf{K}_i = \Phi(\mathbf{X}_i)\Phi(\mathbf{X}_i)^T$, $i = 1, 2, \ldots, S$ with each entry for S different sampling rate. Hence, the measurements with the same sampling rate share the same kernel matrix, which is given as

$$\mathbf{K}_i(h, j) = \left\langle \Phi(\mathbf{x}_{i,h}), \Phi\left(\mathbf{x}_{i,j}\right) \right\rangle = \Phi(\mathbf{x}_{i,h})^T \Phi\left(\mathbf{x}_{i,j}\right), \ i = 1, 2, \ldots, S. \tag{14}$$

Also, $\{\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_{S-1}, \mathbf{k}_S\}$ are the vectors of the kernel matrices $\{\mathbf{K}_1, \mathbf{K}_2, \ldots, \mathbf{K}_{S-1}, \mathbf{K}_S\}$. Using the (9)–(13), the E-step of K-MPPCA can be updated as

$$E\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_i\right) = \mathbf{\Omega}_i^{-1} \sum_{S=1}^{i} \sigma_i^{-2} \mathbf{C}_i^{-1} \mathbf{T}_i^T \mathbf{k}_{i,k}, \ i = 1, 2, \ldots, S \tag{15}$$

$$\mathbf{\Omega}_i^{-1} = \sum_{S=1}^{i} \sigma_S^{-2} \mathbf{C}_S^{-1} \mathbf{T}_S^T \mathbf{K}_S \mathbf{T}_S \mathbf{C}_S^{-1} + \mathbf{I}, \ i = 1, 2, \ldots, S \tag{16}$$

$$E\left(\hat{\mathbf{t}}_k \hat{\mathbf{t}}_k^T \mid \mathbf{V}_i\right) = \mathbf{\Omega}_i^{-1} + E\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_i\right) E^T\left(\hat{\mathbf{t}}_k \mid \mathbf{V}_i\right), \ i = 1, 2, \ldots, S. \tag{17}$$

By introducing the kernel trick, the K-MPPCA model avoids the problem of solving the explicit nonlinear relationships between the process variables and the latent variables. Hence, the posterior expectations of the latent variables can be obtained without loading matrices in the E-

step. Because of this and the assumption that process variables and latent variables are nonlinear, the model K-MPPCA is nonlinear.

Similarly, the kernel trick can also be used to update the model parameters in the M-step, which is given as

$$\hat{\sigma}_i^2 = \frac{1}{M_i K_i}\left\{trace(\mathbf{K}_i) - trace\left(\mathbf{T}_i^T \mathbf{K}_i \mathbf{T}_i \mathbf{C}_i^{-1}\right)\right\},\ \ i = 1,2,\ldots,S \tag{18}$$

in which trace $(\cdot)$ is an operator for the trace value of the matrix. The detailed derivation of the M-step is provided in Appendix A. In the nonlinear EM algorithm, the accurate model parameters can be obtained by iterating E-step and M-step until convergence. It should be noted that the *S*-type measurement data in the multi-rate process were reordered and the corresponding measurement variables were given before the EM algorithm. This is for ease of tagging and coding. The model parameters estimation can still work without this step. In addition, the projected values need to be re-centralized. The specific centralized equations are

$$\bar{\mathbf{K}}_i = \bar{\mathbf{K}}_i - \frac{\mathbf{1}_{n_i}\mathbf{1}_{n_i}^T \mathbf{K}_i}{n_i} - \frac{\mathbf{K}_i \mathbf{1}_{n_i}\mathbf{1}_{n_i}^T}{n_i} + \frac{\mathbf{1}_{n_i}\mathbf{1}_{n_i}^T \mathbf{K}_i \mathbf{1}_{n_i}\mathbf{1}_{n_i}^T}{n_i} \tag{19}$$

where $\mathbf{1}_{n_i} = [1,\ldots,1]^T$, $i = 1,2,\ldots,S$ is the one-column vector of length n_i, which is the size of the data at *ith* sampling rate.

C. Discussions of K-MPPCA

In this subsection, some properties of the proposed K-MPPCA models are further discussed by several remarks.

Remark 1: The proposed K-MPPCA can reduce to KPPCA when S=1, in which the measurements with single sampling rate is used for model training. Also, K-MPPCA can reduce to supervised KPLVR (S=2 and $K1$=$K2$) or semi-supervised KPLVR (S=2 and $K1 \neq K2$). Therefore, these models can be treated as the specific forms for K-MPPCA.

Remark 2: The proposed K-MPPCA model implicitly maps the original multi-rate measurements to a higher dimension. By designing appropriate kernel parameters, the projected values turn to be linearly correlated and the constraint relationship with different sampling rate can be presented by a few linear latent variables.

Remark 3: It is noticed that the K-MPPCA model is totally different from several separated KPPCA models. The latent variables of K-MPPCA are determined by simultaneously considering all the classes of the process variables and the model parameters which are influenced by each other. It can be seen from (15)–(18) that model parameters at different sampling rate are affected by parameters of other sampling rates rather than just themselves.

IV. K-MPPCA Based Fault Detection

It should be noted that the *S*-type measurement data in the multi-rate process were reordered and the corresponding measurement variables were given before the EM algorithm. After the K-MPPCA model is constructed, two commonly used measures T^2 and *SPE* statistics are applied to establish the fault detection scheme. These two statistics are used to monitor the variations of the common principal component space and the residual space, respectively. It is worth noting that the posterior distributions of the latent variables should be separately estimated using different equations when different classes of measurements are collected. Therefore, the calculation method of T^2 and *SPE* statistics of the multi-rate process is different from that in the traditional methods.

When the new samples are collected under different sampling rates, the expected values of the latent variables are calculated separately considering different conditions, which depend on which class of measurements it contains. The estimated mean projections of latent variables are calculated as

$$\mathbf{t}_{i,new} = \mathbf{\Omega}_i^{-1} \sum_{s=1}^{i} \sigma_s^{-2} \mathbf{C}_s^{-1} \mathbf{T}_s^T \mathbf{K}_{new,s}^T,\ \ i = 1,2,...,S \tag{20}$$

where $K_{new,s}$ are the kernel matrices between the test data and normal data. The T^2 statistic can be constructed as

$$T_{i,new}^2 = \mathbf{t}_{i,new}^T \mathrm{var}^{-1}(\mathbf{t}_{i,new}|\Phi(\mathbf{x}_{i,new}))\mathbf{t}_{i,new},\ \ i = 1,2,...,S \tag{21}$$

where $\mathbf{t}_{i,new}$ is the estimated latent variable with the measurements of *ith* classes and $\mathrm{var}^{-1}(\mathbf{t}_{i,new}|\Phi(\mathbf{x}_{i,new}))$ is its corresponding variance and it varies based on the number of the measurements at different sampling time. The confidence limit of T^2 statistics can be gotten by approximating T^2 of normal data by an F-distribution [8]. If T^2 statistics of test data exceed the confidence limit, the corresponding sample is thought as a fault.

In addition, *SPE* statistics can also be constructed based on the prediction errors of the model. Considering that the test data may have multiple sampling rates, it must design S type of *SPE* statistics for S sampling rates, which is given as

$$\begin{aligned} SPE_{i,new} &= \mathbf{e}_{i,new}^T \mathbf{e}_{i,new} \\ &= [\Phi(\mathbf{x}_{i,new}) - \Phi(\hat{\mathbf{x}}_{i,new})]^T [\Phi(\mathbf{x}_{i,new}) - \Phi(\hat{\mathbf{x}}_{i,new})] \\ &\quad i = 1,2,...,S. \end{aligned} \tag{22}$$

Since the projected values $\Phi(\mathbf{x}_{i,new})$ cannot be directly obtained, the kernel function is also used to eliminate the inner product of the high-dimensional mapping set and they are given as

$$\begin{aligned} SPE_{i,new} &= \mathbf{K}_{i,new} - 2\mathbf{K}_{i,new}\mathbf{T}_i\mathbf{C}_i^{-1}\mathbf{t}_{i,new} \\ &\quad + (\mathbf{T}_i\mathbf{C}_i^{-1}\mathbf{t}_{i,new})^T \mathbf{K}_i \mathbf{T}_i \mathbf{C}_i^{-1}\mathbf{t}_{i,new} \\ &\quad i = 1,2,...,S. \end{aligned} \tag{23}$$

The detailed derivation of $SPE_{i,new}$ are provided in Appendix B. It is noticed that the SPE statistics are separately constructed for different sampling rates. Hence, they can reflect the fluctuations of process variables for each sampling rate in the residual space and indicate the particular fault. The confidence line of the SPE statistics can also be estimated by an χ^2 distributed approximation: $SPE \sim g \cdot \chi_h^2$, in which $\boldsymbol{g}$ and $\boldsymbol{h}$ are the parameters of χ^2 distribution and they are given as [39]

$$\begin{aligned} g \cdot h &= mean(SPE_{normal}) \\ 2g^2 h &= \mathrm{var}(SPE_{normal}). \end{aligned} \tag{24}$$

V. Case Study

In this section, a numerical example and an actual pre-

decarbonization unit in the ammonia synthesis process are tested to evaluate the process modeling and monitoring performance of the proposed method. The first one is a hypothetical example, as it is easier to understand why the proposed K-MPPCA outperforms the traditional models. Then, we applied the proposed method to a pre-decarbonization unit data in the ammonia synthesis process found in the chemical industry. In both cases, the process variables are highly nonlinearly correlated, in which several nonlinear process monitoring methods are compared, especially when the process sampling rates are different.

A. Numerical Example

In order to demonstrate the effectiveness of the proposed algorithm, a multi-sampling rate process is designed in this section. There are 8 variables in the system, which are obtained by a nonlinear combination of 6 latent variables. The process is given as follows:

$$\begin{aligned}
x_1 &= 0.1t_1 + t_2 \div \mathrm{sqr}t({t_6}^2 + {t_2}^2) + 0.3t_4 \\
x_2 &= 0.1t_1 \times t_2 + t_2 \div \mathrm{sqr}t({t_6}^2 + {t_2}^2) \\
x_3 &= t_2 \times (\cos {t_3}^3) + 0.1\exp(\sin(t_2)) \\
x_4 &= \sin(t_1)^3 + 0.2 \times \log(2 + \cos(t_4)) \\
x_5 &= \exp(\cos(t_2)) + 3 \times \log(30 - \sin(t_5)) + t_6 \\
x_6 &= 0.1t_4 + \exp(t_3 + t_1) \\
x_7 &= 0.5t_5 + t_1 + \sin(t_2) \\
x_8 &= \exp(t_1) + \log(\sin(t_6) + 5) + t_3
\end{aligned} \tag{25}$$

where $\{x_1, x_2, \ldots, x_8\}$ are process variables and $\{t_1, \ldots, t_6\}$ are latent variables. Besides, some small disturbances are added to simulate the measurement noises, which are assumed to follow Gaussian distribution as $N\left(\mathbf{0}, 0.1^2\mathbf{I}\right)$. In this process, these variables are collected in different sampling rates, where $\{x_1, x_2, \ldots, x_5\}$ are sampled once a minute, $\{x_6, x_7\}$ are sampled every two minutes, and x_8 is collected every ten minutes. In total, these variables are generated within a period of 2000 minutes, which indicates that 2000 samples of $\{x_1, x_2, \ldots, x_5\}$, 1000 samples of $\{x_6, x_7\}$, and 200 x_8 have been collected. For comparison, KPPCA and Q-KPCA are also used for nonlinear process modeling and monitoring [29], [31]. Since KPPCA is based on the single sampling rate data, there are two types of data pre-processing methods used in this paper. The first one only uses the measurements with the highest sampling rate, which means that only $\{x_1, x_2, \ldots, x_5\}$ are used for the model training. This model is named as KPPCA_1. Another choice is to utilize the samples where all the variables are collected at that sampling interval. Hence, only 200 samples are obtained, which are based on the lowest sampling rate. This model is named KPPCA_2. The dimension of the feature space of K-MPPCA is selected as 10. The widely used radial basis function (RBF) is selected as the kernel function. The kernel parameters are chosen as 450, 30, and 0.1 for different sampling rates. To be fair, the component numbers of KPPCA_1 and KPPCA_2 are both set as 10 and their kernel parameters are both selected as 450: the same as that in the K-MPPCA model with the highest sampling rate. The choice of the kernel parameters is still an open question in kernel learning methods. In this paper, the cross-validation is used for model selection and it is obtained that these three methods can achieve the best results under the same kernel parameters. Moreover, the fault detection effect of the model does not change greatly when the kernel parameters are kept within a certain range. For the Q-KPCA model, in this case, $\{x_1, x_2, \ldots, x_5\}$ are the process variables and x_8 is the key quality variable in Q-KPCA. Again, only the data at the full sampling interval are utilized. The kernel parameter is set to 500 according to the cross-validation method. The dimensions of the latent variables in quality-related space and quality-uncorrelated space are set to be 1 and 50, respectively. In Q-KPCA, T_y^2 and T_{o}^2 statistics are used to monitor the changes of these two spaces respectively.

For fault detection purposes, another set of samples with a period of 2000 minutes are generated, in which five kinds of faults are introduced from the 1001st minute to evaluate the fault detection performance for different models. Like most fault detection methods, both T^2 and SPE statistics are used to measure the effect of fault detection in this case. The calculation methods of the control limit of two statistics have been mentioned in the previous section. If the value of the statistic is greater than the control limit, it is considered a fault. The detailed fault descriptions are given in Table I.

TABLE I
FAULT DESCRIPTION IN THE NUMERICAL EXAMPLE

Fault	Fault variable	Fault type	Time
1	1st latent variable	Step change by 1	1001–2000
2	3rd latent variable	Ramp change by 0.005×t	1001–2000
3	4th process variable	Additive random fault with $N(0,5)$	1001–2000
4	8th process variable	Structure change by log to exp	1001–2000
5	7th process variable	Step change by 1	1001–2000

In this paper, the false alarm rates and missing detection rates of faults are used to qualify the process monitoring results. Among them, false alarm rates refer to the proportion of normal samples identified as fault samples, while missing detection rates refer to the proportion of fault samples detected as normal samples. Given a significance level of 99 percent, the detailed fault detection results are given in Table II.

The lower false alarm and missing detection rate, the better the model performance. The results with better detection performance are highlighted in bold and the fault 0 data is the false alarm rate of these three models.

As shown in Table II, the fault detection performance of the K-MPPCA model is better than the other two models for most cases. The main reason is that the proposed K-MPPCA model has made use of all the collected data to extract the nonlinear correlations among different sampling rates. For comparison, KPPCA_1 does not consider the low sampling data and KPPCA_2 has neglected too many useful fast sampling data. The Q-KPCA model only shows good detection effect on Fault 3. This shows that the full utilization of multi-sampling rate data can effectively improve the performance of nonlinear process monitoring. Take Fault 1 for example: the false alarm rate of the T^2 and *SPE* statistics for the K-MPPCA model is lower than the other three alternatives. The detailed fault

TABLE II
MONITORING RESULTS IN THE NUMERICAL EXAMPLE

Fault	K-MPPCA				$KPPCA_1$		$KPPCA_2$		Q-KPCA	
	T^2	SPE_1	SPE_2	SPE_3	T^2	SPE	T^2	SPE	T^2_y	T^2_o
0	**0.067**	0.189	**0.038**	0.2	0.08	0.281	0.03	0.16	0.01	0.04
1	0.134	**0.157**	**0.242**	**0.01**	0.08	0.293	0.97	0.55	0.140	0.970
2	0.468	0.586	**0.138**	**0.38**	0.497	0.436	0.15	1	0.950	1
3	1	**0**	0.804	0.98	0.992	0	1	0	0	0.010
4	0.941	0.749	0.94	**0.35**	0.93	0.696	1	0	1	1
5	0.915	0.99	**0.096**	0.8	0.919	0.929	0	0.1	1	0.970

detection results of Fault 1 using different models are shown in Fig. 1.

In addition, the *SPE* statistics are separately built for different sampling rates for the K-MPPCA based fault detection strategy since the measurements with multiple sampling rates are usually collected from different sources, such as process and quality variables. Hence, such a design will be helpful for further fault identification. For example, Fault 4 and Fault 5 are about the process variables x_7 and x_8. However, $KPPCA_1$ did not use variables x_7 and x_8 during modeling. Hence, the fault about x_7 and x_8 will not be detected using $KPPCA_1$.

In practice, such faults are similar to the sensor faults, which often exist and whose sampling rate of fault variables is sometimes low. If only the highest-sampling-rate data are used for modeling like $KPPCA_1$, such faults will not be detected totally, which will cause great damage. Hence, this situation also highlights the advantages of our proposed K-MPPCA over $KPPCA_1$ which are shown in Fig. 2.

In addition, since different *SPE* statistics are established for variables with different sampling rates, the K-MPPCA model can indicate the specified sampling rate according to the *SPE* statistics. This is also helpful for further fault identification. For $KPPCA_2$, however, it only contained part of sampled data so that the established model is not accurate enough and the false alarm rate is too high. By contrast, the T^2 and statistics in the K-MPPCA model can accurately detect the potential faults with a low false alarm rate. At the same time, only the *SPE* statistic at the second sampling rate has exceeded the control limit, which can be used to determine the range of variables that are causing the fault. Furthermore, it is useful for further fault identification. The detailed detection result of Fault 5 is shown in Fig. 2.

B. Pre-Decarburization Unit

The ammonia synthesis process is quite common in modern chemical industries and its product, NH_3, is also the basic material for many processes, such as the urea synthesis. Pre-decarburization is a crucial unit to the process and its main function is to eliminate carbon dioxide (CO2) in the original process gas (PG) as much as possible. The flowchart of the pre-decarburization unit with all instruments is given in Fig. 3.

The process description and the sampling rates of measurements are tabulated in Table III.

From the DCS database in the ammonia synthesis process, nine process variables under the normal operation condition within a period of 3000 min are collected. As shown in

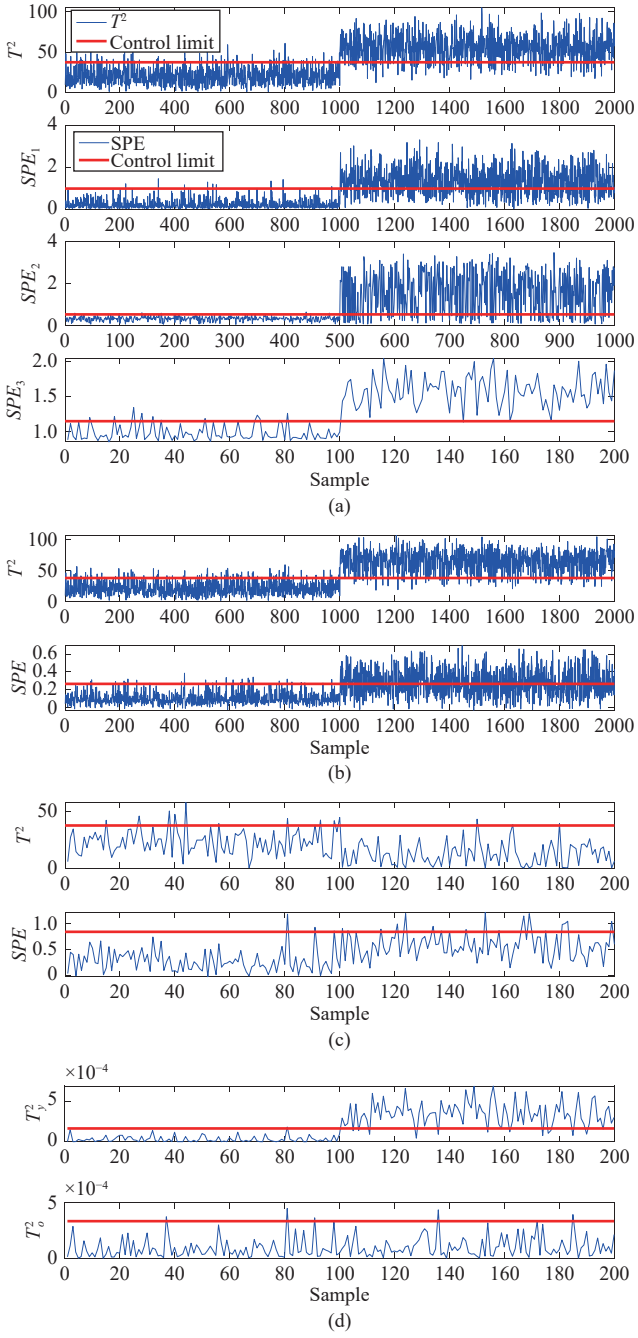

Fig. 1. Monitoring results of Fault 1 in the numerical example by (a) K-MPPCA, (b) $KPPCA_1$, (c) $KPPCA_2$, (d) Q-KPCA.

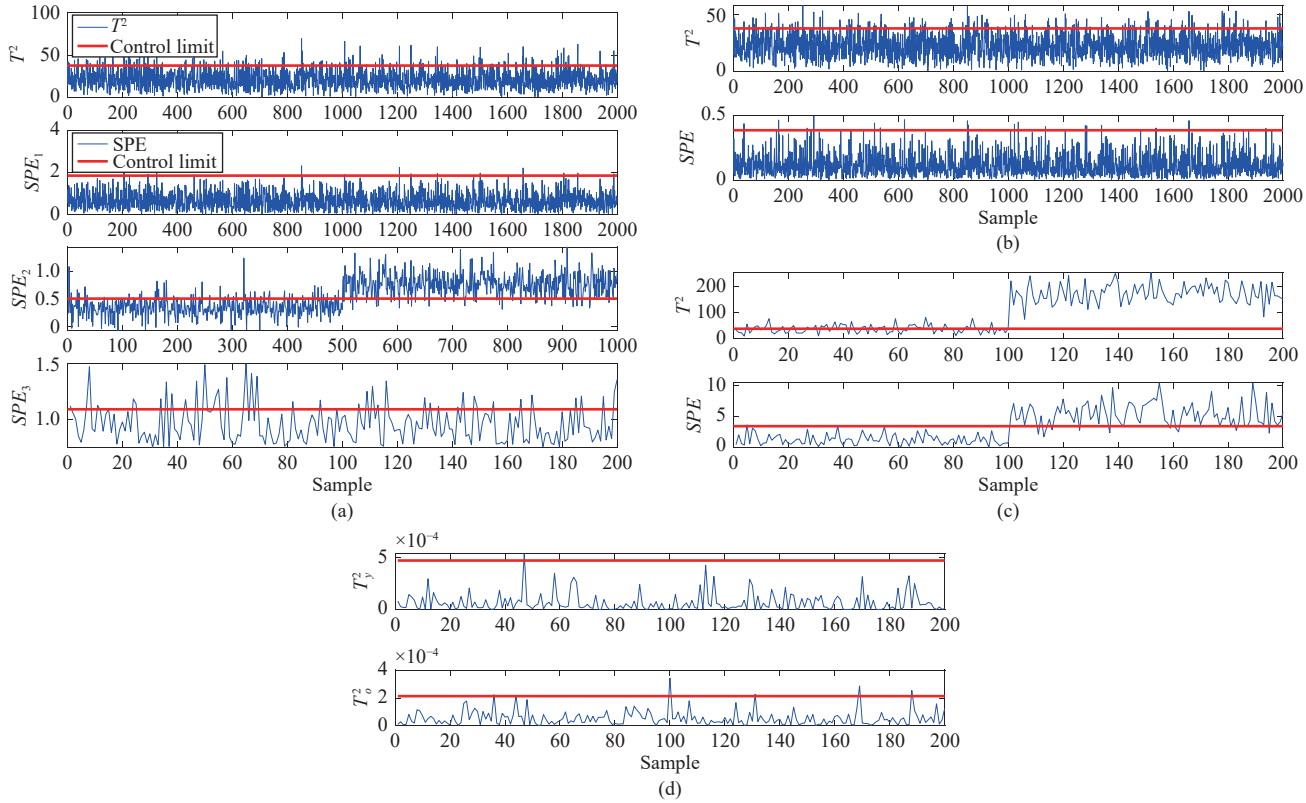

Fig. 2. Monitoring results of Fault 5 in the numerical example by (a) K-MPPCA, (b) $KPPCA_1$, (c) $KPPCA_2$, (d) Q-KPCA.

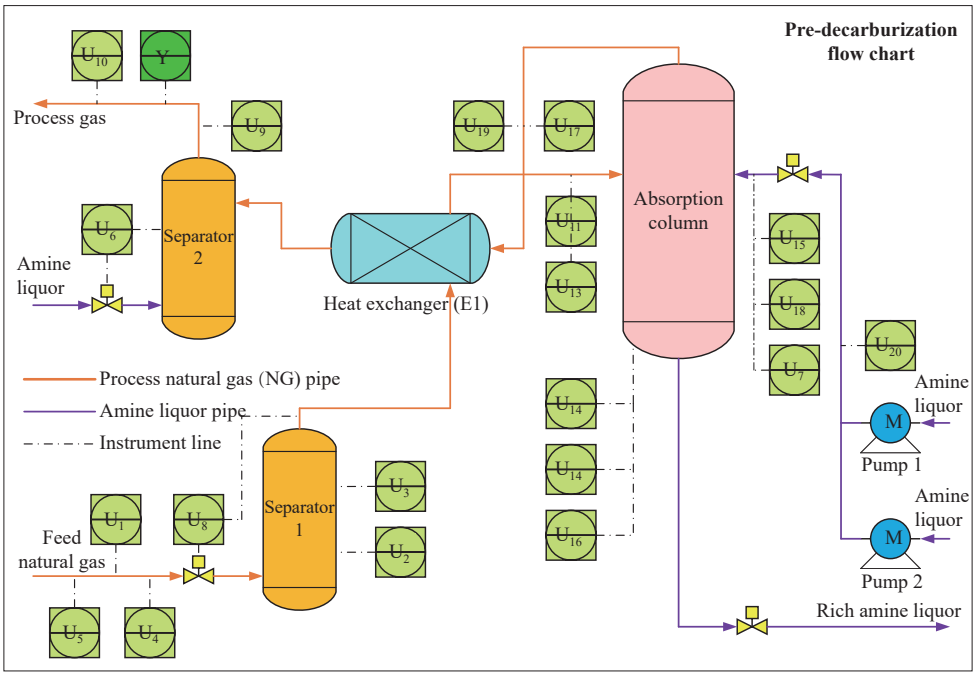

Fig. 3. Flow chart of the pre-decarbonization unit.

Table III, there are three kinds of sampling rates in these measurements, in which 4 of them are sampled per minute and the other variables are sampled every 2 min or 10 min. Hence, 3000 XMEAS (1)–(4), 1500 XMEAS (5)–(7) and 300 XMEAS (8)–(9) are collected in total. As in the last case, the proposed K-MPPCA, $KPPCA_1$, $KPPCA_2$, and Q-KPCA models are built based on these normal datasets. Similarly, the $KPPCA_1$ model only uses the data with the highest sampling rate for modeling. Although the sampling number is as large as the K-MPPCA model, other low sampled variables are not utilized. $KPPCA_2$ model only selects the data at the sampling interval when all variables have been recorded, which indicates that only 300 samples are used. The Q-KPCA model uses XMEAS (1)–(4) as process variables and XMEAS (9) as the quality variable. Similar to $KPPCA_2$, only 300 samples which contain all the measurements are used for training. In

TABLE III
PROCESS DESCRIPTION AND SAMPLING RATES OF THE PRE-DECARBURIZED UNIT

Number	Tags	Descriptions	Sampling rate
XMEAS (1)	U12	Level A of 15-C001	1 min
XMEAS (2)	U14	Level B of 15-C001	
XMEAS (3)	U16	Level C of 15-C001	
XMEAS (4)	U20	Level of recovery column	
XMEAS (5)	U1	Flow-rate of input natural gas	2 min
XMEAS (6)	U2	Level of 15-F001	
XMEAS (7)	U3	Pressure differential of 15-F001	
XMEAS (8)	U5	Temperature of input natural gas	10 min
XMEAS (9)	U9	Temperature of outlet gas at 15-F002	

this case, the traditional Gaussian kernel is still selected as the kernel method and the kernel parameter of K-MPPCA is set to be 1000 while the kernel parameters of two KPPCA models and the Q-KPCA model are set to be 800 according to the tuning result. The dimension of all feature spaces of both the K-MPPCA and KPPCA model are set to 20. According to the fluctuation of the raw materials, the DCS database has recorded two types of faults. To measure the effectiveness of fault detection, another 3000 min of process variables were collected as the test data and the faults are started at the 1501st min. The false alarm rate and missing detection rate are used to compare the detection performance of different models. The results for both faults are given in Table IV.

It can be found that the fault detection performance of the K-MPPCA model is improved compared with alternatives, as in the last case. For both faults, the abnormal condition can be opportunely detected with the low false alarm rates. In contrast, the false alarm rates of the two types of KPPCA models are much higher, which indicates the control limits of these two models are wrong and the corresponding detection results are invalid. Especially, the missing detection rates of T^2 and *SPE* statistics in Fault 1 are all close to zero with the normal false alarm rates. However, the detection performance of the normal test data using KPPCA_1 and KPPCA_2 are both higher so that the corresponding fault detection results invalid. The fault detection performance of the Q-KPCA model performs better than KPPCA model due to its functionality with nonlinear models. However, K-MPPCA is still superior since it can deal with the multi-rate modeling problems. The detailed process monitoring result of Fault 1 is shown in Fig. 4.

VI. CONCLUSIONS

In this paper, a kernel generalized MPPCA model has been proposed for nonlinear process modeling and monitoring

TABLE IV
MONITORING RESULTS IN THE PRE-DECARBONIZATION UNIT

Fault	KMPPCA				KPPCA_1		KPPCA_2		Q-KPCA	
	T^2	SPE_1	SPE_2	SPE_3	T^2	*SPE*	T^2	*SPE*	T_y^2	T_o^2
Normal	**0.0480**	**0.0567**	**0.0107**	**0.0133**	0.58	0.008	0.9933	0.6733	0.0200	0.3467
Fault 1	0.0120	**0**	**0**	**0**	0	0.1020	0	0.0067	0.0800	0.0200
Fault 2	**0**	**0.0307**	**0**	**0**	0.0020	0.9207	0	0	0.4866	0.1666

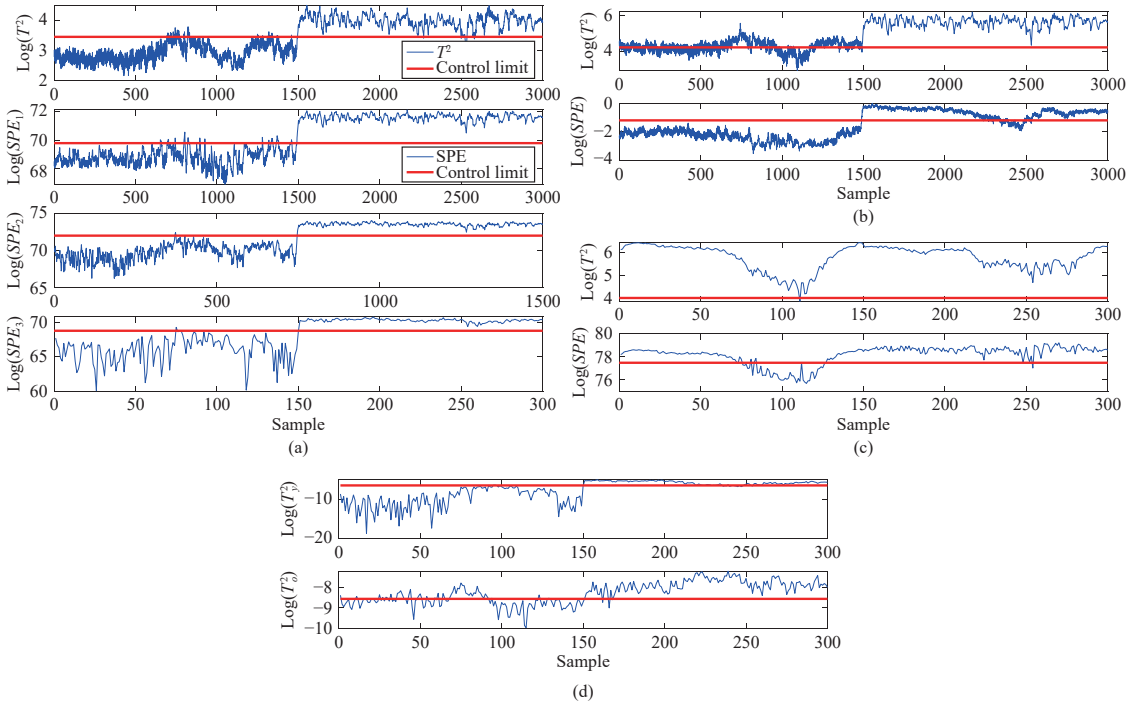

Fig. 4. Monitoring results of Fault 1 in the pre-decarbonization unit by (a) K-MPPCA, (b) KPPCA_1, (c) KPPCA_2, (d) Q-KPCA.

purpose. With the introduction of the kernel trick for process variables with different sampling rates, a nonlinear multi-rate probabilistic model is derived to utilize all the measurements from different sampling scales. Also, the corresponding fault detection approach is developed based on the proposed K-MPPCA. The case study of numerical example and an actual industrial process example prove the feasibility of the proposed method. In our future studies, more nonlinear modeling technologies like the deep learning models will be explored for multi-rate process monitoring purpose, in which the nonlinear modeling performance for the measurements with different sampling rates are desired to be further improved. At the same time, more complex data characteristics like the auto-correlations will be considered simultaneously.

APPENDIX A
DETAIL DERIVATION OF THE M-STEP

As we defined above, the parameters estimation equations are given as follows:

$$\hat{\sigma}_i^2=\frac{1}{M_iK_i}\left\{\begin{array}{l}\displaystyle\sum_{k=1}^{K_i}\boldsymbol{\Phi}(\mathbf{x}_{i,k})^T\boldsymbol{\Phi}(\mathbf{x}_{i,k})\\ \quad -2\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_S)\hat{\mathbf{P}}_i^T\boldsymbol{\Phi}(\mathbf{x}_{ik})\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_{S-1})\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{ik})\\ \quad+\cdots+\displaystyle\sum_{k=K_{i+1}+1}^{K_i}E^T(\mathbf{t}_k|\mathbf{V}_i)\hat{\mathbf{P}}_i^T\boldsymbol{\Phi}(\mathbf{x}_{ik})\end{array}\right)\\ \quad +trace\left(\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_S)\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_{S-1})\\ \quad+\cdots+\displaystyle\sum_{k=K_i+1}^{K_i}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_i)\end{array}\right)\hat{\mathbf{P}}_i^T\hat{\mathbf{P}}_i\right)\end{array}\right\}$$

$$i=1,2,\ldots,S. \tag{26}$$

With the introduction of the kernel tricks, (8) can be all rewritten to (18). Hence, we take the parameter estimation equation of $\hat{\sigma}_1^2$ as an example:

$$\hat{\sigma}_1^2=\frac{1}{M_1K_1}\left\{\begin{array}{l}\displaystyle\sum_{k=1}^{K_1}\boldsymbol{\Phi}(\mathbf{x}_{1k})^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad -2\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_S)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_{S-1})\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E^T(\mathbf{t}_k|\mathbf{V}_1)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\end{array}\right)\end{array}\right\}$$

$$\hat{\sigma}_1^2=\frac{1}{M_1K_1}\left\{+trace\left(\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_S)\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_{S-1})\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_1)\end{array}\right)\hat{\mathbf{P}}_1^T\hat{\mathbf{P}}_1\right)\right\}$$

$$=\frac{1}{M_1K_1}\left\{\begin{array}{l}trace(\mathbf{K}_1)+trace\\ \left(\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_S)\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_{S-1})\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E(\hat{\mathbf{t}}_k\hat{\mathbf{t}}_k^T|\mathbf{V}_1)\end{array}\right)\hat{\mathbf{P}}_1^T\hat{\mathbf{P}}_1\right)\\ -2\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_S)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_{S-1})\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E^T(\mathbf{t}_k|\mathbf{V}_1)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\end{array}\right)\end{array}\right\}$$

$$=\frac{1}{M_1K_1}\left\{\begin{array}{l}trace(\mathbf{K}_1)+trace\left(\mathbf{C}_1\hat{\mathbf{P}}_1^T\hat{\mathbf{P}}_1\right)\\ -2\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_S)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_{S-1})\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E^T(\mathbf{t}_k|\mathbf{V}_1)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\end{array}\right)\end{array}\right\}$$

$$=\frac{1}{M_1K_1}\left\{\begin{array}{l}trace(\mathbf{K}_1)+trace\left(\mathbf{T}_1{}^T\mathbf{K}_1\mathbf{T}_1\mathbf{C}_1{}^{-1}\right)\\ -2\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_S)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_{S-1})\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E^T(\mathbf{t}_k|\mathbf{V}_1)\hat{\mathbf{P}}_1^T\boldsymbol{\Phi}(\mathbf{x}_{1k})\end{array}\right)\end{array}\right\}$$

$$=\frac{1}{M_1K_1}\left\{\begin{array}{l}trace(\mathbf{K}_1)+trace\left(\mathbf{T}_1{}^T\mathbf{K}_1\mathbf{T}_1\mathbf{C}_1{}^{-1}\right)\\ -2\left(\begin{array}{l}\displaystyle\sum_{k=1}^{K_S}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_S)\mathbf{C}_1{}^{-1}\mathbf{T}_1{}^T\mathbf{k}_{1k}\\ \quad+\displaystyle\sum_{k=K_S+1}^{K_{S-1}}E^T(\hat{\mathbf{t}}_k|\mathbf{V}_{S-1})\mathbf{C}_1{}^{-1}\mathbf{T}_1{}^T\mathbf{k}_{1k}\\ \quad+\cdots+\displaystyle\sum_{k=K_2+1}^{K_1}E^T(\mathbf{t}_k|\mathbf{V}_1)\mathbf{C}_1{}^{-1}\mathbf{T}_1{}^T\mathbf{k}_{1k}\end{array}\right)\end{array}\right\}$$

$$= \frac{1}{M_1 K_1} \left\{ \begin{array}{l} trace(\mathbf{K}_1) + trace\left(\mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1}\right) \\ -2trace\left(\mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1}\right) \end{array} \right\}$$

$$= \frac{1}{M_1 K_1} \left\{ trace(\mathbf{K}_1) - trace\left(\mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1}\right) \right\}. \quad (27)$$

Based on the derivation in (1), the estimation equations for other parameters can be rewritten in a similar way as

$$\hat{\sigma}_1^2 = \frac{1}{M_1 K_1} \left\{ trace(\mathbf{K}_1) - trace\left(\mathbf{T}_1^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1^{-1}\right) \right\}$$

$$\hat{\sigma}_2^2 = \frac{1}{M_2 K_2} \left\{ trace(\mathbf{K}_2) - trace\left(\mathbf{T}_2^T \mathbf{K}_2 \mathbf{T}_2 \mathbf{C}_2^{-1}\right) \right\}$$

$$\vdots$$

$$\hat{\sigma}_{S-1}^2 = \frac{1}{M_{S-1} K_{S-1}} \left\{ trace(\mathbf{K}_{S-1}) - trace\left(\mathbf{T}_{S-1}^T \mathbf{K}_{S-1} \mathbf{T}_{S-1} \mathbf{C}_{S-1}^{-1}\right) \right\}$$

$$\hat{\sigma}_S^2 = \frac{1}{M_S K_S} \left\{ trace(\mathbf{K}_S) - trace\left(\mathbf{T}_S^T \mathbf{K}_S \mathbf{T}_S \mathbf{C}_S^{-1}\right) \right\}. \quad (28)$$

APPENDIX B
DETAIL DERIVATION OF $\{SPE_{1,new}, SPE_{2,new}, ..., SPE_{S-1,new}, SPE_{S,new}\}$

In (22), the projection of the process variables can not be directly obtained. Hence, (22) can be rewritten as follow by introducing the kernel matrix. Here we take the equation of SPE_1 as the example:

$$\begin{aligned} SPE_{1,new} &= \mathbf{e}_1^T \mathbf{e}_1 = [\Phi(\mathbf{x}_1) - \Phi(\hat{\mathbf{x}}_1)]^T [\Phi(\mathbf{x}_1) - \Phi(\hat{\mathbf{x}}_1)] \\ &= \Phi(\mathbf{x}_1)^T \Phi(\mathbf{x}_1) - \Phi(\hat{\mathbf{x}}_1)^T \Phi(\mathbf{x}_1) \\ &\quad - \Phi(\mathbf{x}_1)^T \Phi(\hat{\mathbf{x}}_1) + \Phi(\hat{\mathbf{x}}_1)^T \Phi(\hat{\mathbf{x}}_1) \\ &= \Phi(\mathbf{x}_1)^T \Phi(\mathbf{x}_1) - \Phi(\hat{\mathbf{x}}_1)^T \Phi(\mathbf{x}_1) \\ &\quad - \Phi(\mathbf{x}_1)^T \Phi(\hat{\mathbf{x}}_1) + \Phi(\hat{\mathbf{x}}_1)^T \Phi(\hat{\mathbf{x}}_1) \\ &= \mathbf{K}_{1,new} - (\mathbf{P}_1 \mathbf{t}_{1,new})^T \Phi(\mathbf{x}_1) - \Phi(\mathbf{x}_1)^T \mathbf{P}_1 \mathbf{t}_{1,new} \\ &\quad + (\mathbf{P}_1 \mathbf{t}_{1,new})^T \mathbf{P}_1 \mathbf{t}_{1,new} \\ &= \mathbf{K}_{1,new} - \mathbf{t}_{1,new}^T \mathbf{P}_1^T \Phi(\mathbf{x}_1) - (\mathbf{P}_1^T \Phi(\mathbf{x}_1))^T \mathbf{t}_{1,new} \\ &\quad + \mathbf{t}_{1,new}^T \mathbf{P}_1^T \mathbf{P}_1 \mathbf{t}_{1,new} \\ &= \mathbf{K}_{1,new} - \mathbf{t}_{1,new}^T \mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \Phi(\mathbf{X}_1)^T \Phi(\mathbf{x}_1) \\ &\quad - (\mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \Phi(\mathbf{X}_1)^T \Phi(\mathbf{x}_1))^T \mathbf{t}_{1,new} \\ &\quad + \mathbf{t}_{1,new}^T \mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1} \mathbf{t}_{1,new} \\ &= \mathbf{K}_{1,new} - \mathbf{t}_{1,new}^T \mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_{new,1} \\ &\quad - (\mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_{new,1})^T \mathbf{t}_{1,new} \\ &\quad + \mathbf{t}_{1,new}^T \mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1} \mathbf{t}_{1,new} \\ &= \mathbf{K}_{1,new} - 2(\mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_{new,1})^T \mathbf{t}_{1,new} \\ &\quad + \mathbf{t}_{1,new}^T \mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1} \mathbf{t}_{1,new} \\ &= \mathbf{K}_{1,new} - 2\mathbf{K}_{new,1} \mathbf{T}_1 \mathbf{C}_1^{-1} \mathbf{t}_{1,new} \\ &\quad + \mathbf{t}_{1,new}^T \mathbf{C}_1{}^{-1} \mathbf{T}_1{}^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1} \mathbf{t}_{1,new} \\ &= \mathbf{K}_{1,new} - 2\mathbf{K}_{new,1} \mathbf{T}_1 \mathbf{C}_1^{-1} \mathbf{t}_{1,new} \\ &\quad + (\mathbf{T}_1 \mathbf{C}_1{}^{-1} \mathbf{t}_{1,new})^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1{}^{-1} \mathbf{t}_{1,new} \end{aligned} \quad (29)$$

where $\mathbf{K}_{1,new}$ is the kernel matrix of the test data. And $\mathbf{K}_{new,1}$ is the kernel matrix of the test data and the normal data. According to the equation above, the whole SPE statistics can be calculated as

$$\begin{aligned} SPE_{1,new} &= \mathbf{K}_{1,new} - 2\mathbf{K}_{new,1} \mathbf{T}_1 \mathbf{C}_1^{-1} \mathbf{t}_{1,new} \\ &\quad + (\mathbf{T}_1 \mathbf{C}_1^{-1} \mathbf{t}_{1,new})^T \mathbf{K}_1 \mathbf{T}_1 \mathbf{C}_1^{-1} \mathbf{t}_{1,new} \\ SPE_{2,new} &= \mathbf{K}_{2,new} - 2\mathbf{K}_{new,2} \mathbf{T}_2 \mathbf{C}_2^{-1} \mathbf{t}_{2,new} \\ &\quad + (\mathbf{T}_2 \mathbf{C}_2^{-1} \mathbf{t}_{2,new})^T \mathbf{K}_2 \mathbf{T}_2 \mathbf{C}_2^{-1} \mathbf{t}_{2,new} \\ &\vdots \\ SPE_{S-1,new} &= \mathbf{K}_{S-1,new} - 2\mathbf{K}_{new,S-1} \mathbf{T}_{S-1} \mathbf{C}_{S-1}^{-1} \mathbf{t}_{S-1,new} \\ &\quad + (\mathbf{T}_{S-1} \mathbf{C}_{S-1}^{-1} \mathbf{t}_{S-1,new})^T \mathbf{K}_{S-1} \mathbf{T}_{S-1} \mathbf{C}_{S-1}^{-1} \mathbf{t}_{S-1,new} \\ SPE_{S,new} &= \mathbf{K}_{S,new} - 2\mathbf{K}_{newS} \mathbf{T}_S \mathbf{C}_S^{-1} \mathbf{t}_{Snew} \\ &\quad + (\mathbf{T}_S \mathbf{C}_S^{-1} \mathbf{t}_{Snew})^T \mathbf{K}_S \mathbf{T}_S \mathbf{C}_S^{-1} \mathbf{t}_{Snew}. \end{aligned} \quad (30)$$

REFERENCES

[1] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control*, vol. 36, no. 2, pp. 220–234, 2012.

[2] Q. Liu, T. Y. Chai, S. J. Qin, and J. Zhao, "Progress of data-driven and knowledge-driven process monitoring and fault diagnosis for industry process," *Control and Decision*, vol. 25, no. 6, pp. 801–807, 2010.

[3] Z. Ge, Z. Song, and F. Gao, "Review of Recent Research on Data-Based Process Monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.

[4] 4. S Yin, S. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Industrial Electronics*, vol. 61, no. 11, pp. 6418–6428, 2014.

[5] W. Li, H. Yue, S, Valle-Cervantes, and S. J. Qin, "Recursive PCA for adaptive process monitoring," *Journal of Process Control*, vol. 10, no. 5, pp. 471–486, 2000.

[6] S. Yin, S. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.

[7] W. Woodall and D. C. Montgomery, "Some current directions in the theory and application of statistical process monitoring," *Journal of Quality Technology*, vol. 46, no. 1, pp. 78–94, 2014.

[8] S. J. Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8–9, pp. 480–502, 2003.

[9] J. F. Macgregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE Journal*, vol. 40, no. 5, pp. 826–838, 1994.

[10] J. Chen and K. C. Liu, "On-line batch process monitoring using dynamic PCA and dynamic PLS models," *Chemical Engineering Science*, vol. 57, no. 1, pp. 63–75, 2002.

[11] G. Li, S. J. Qin, and D. Zhou, "Geometric properties of partial least squares for process monitoring," *Automatica*, vol. 46, no. 1, pp. 204–210, 2010.

[12] J. Huang, S. Yan, and X. Yan, "Robust chemical process monitoring based on CDC-MVT-PCA eliminating outliers and optimally selecting principal component," *Canadian Journal of Chemical Engineering*, vol. 97, no. 6, pp. 1848–1857, 2019.

[13] Z. Ge, "Process data analytics via probabilistic latent variable models: A tutorial review," *Industrial & Engineering Chemistry Research*, vol. 57, no. 38, pp. 12646–12661, 2018.

[14] Z. Ge and Z. Song, "Maximum-likelihood mixture factor analysis model and its application for process monitoring," *Chemometrics & Intelligent Laboratory Systems*, vol. 102, no. 1, pp. 53–61, 2010.

[15] J. Zheng, L. Zhou, Z. Ge, and Z. Song, "Switching autoregressive dynamic latent variable model for fault detection in multimode

processes. Data Driven Control and Learning Systems", in *Proc. 6th Data Driven Control and Learning Systems (DDCLS)*, 2017, pp. 617–622.

[16] l. Zhou, J. Zheng, Z. Ge, Z. Song, and S. Shan, "Multimode process monitoring based on switching autoregressive dynamic latent variable model," *IEEE Trans. Industrial Electronics*, vol. 65, no. 10, pp. 8184–8194, 2018.

[17] C. Chen, Z. Liu, W. H. Lin, S. Li and K. Wang, "Distributed modeling in a mapreduce framework for data-driven traffic flow forecasting," *IEEE Trans. Intelligent Transportation Systems*, vol. 14, no. 1, pp. 22–33, 2013.

[18] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: the role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[19] S. Yin and O. Kaynak, "Big data for moden industy: Challenges and trends," *IEEE Trans. Industrial Electronics*, vol. 103, no. 2, pp. 143–146, 2015.

[20] N. Lu, Y. Yang, F. Ga, and F. Wang, "Multirate dynamic inferential modeling for multivariable processes," *Chemical Engineering Science*, vol. 59, no. 4, pp. 855–864, 2004.

[21] O. Marjanovic, B. Lennox, D. Sandoz, K. Smith, and M. Crofts, "Real-time monitoring of an industrial batch process," *Computers & Chemical Engineering*, vol. 30, no. 10–12, pp. 1476–1481, 2006.

[22] D. Li, S.L Shah, and T. Chen, "Identification of fast-rate models from multirate data," *Int. Journal of Control*, vol. 74, no. 7, pp. 680–689, 2001.

[23] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), vol. 61, no. 3, pp. 611–622, 1999.

[24] Z. Ge, B. Huang, and Z. Song, "Mixture semi-supervised principal component regression model and soft sensor application," *American Institute of Chemical Engineers Journal*, vol. 60, no. 2, pp. 533–545, 2014.

[25] L. Zhou, J. Chen, Z. Song, Z. Ge, and A. Miao, "Probabilistic latent variable regression model for process-quality monitoring," *Chemical Engineering Science*, vol. 116, pp. 296–305, 2014.

[26] L. Zhou, J. Chen, J. Jie, and Z. Song, "Multiple probability principal component analysis for process monitoring with multi-rate measurements," *Journal of the Taiwan Institute of Chemical Engineers*, vol. 96, pp. 18–28, 2018.

[27] L. Zhou, J. Chen, J. Jie, and Z. Song, "Multirate Factor Analysis Models for Fault Detection in Multirate Processes," *IEEE Trans. Industrial Informatics*, vol. 15, no. 7, pp. 4076–4085, 2019.

[28] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, USA: Springer, 2006.

[29] Z. Ge, C. Yang, and Z. Song, "Improved kernel-PCA based monitoring approach for nonlinear processes," *Chemical Engineering Science*, vol. 64, pp. 2245–2255, 2009.

[30] Z. Ge and Z. Song, "Kernel Generalization of PPCA for Nonlinear Probabilistic Monitoring," *Industrial & Engineering Chemistry Research*, vol. 49, pp. 11832–11836, 2010.

[31] L. Zhou, Z. Song, B. Hou, and Z. Fei, "Robust semi-supervised modeling method and its application to fault detection in chemical processes," *Journal of Chemical Industry and Engineering* (*China*), vol. 68, no. 3, pp. 1109–1115, 2017.

[32] J. Huang and X. Yan, "Quality-Driven Principal Component Analysis Combined with Kernel Least Squares for Statistical Process Monitoring," *IEEE Trans. Control Systems Technology*, vol. 27, no. 6, pp. 2688–2695, 2019.

[33] Z. Li and X. Yan, "Ensemble model of wastewater treatment plant based on rich diversity of principal component determining by genetic algorithm for status monitoring," *Control Engineering Practice*, vol. 88, pp. 38–51, 2019.

[34] Z. Li and X. Yan, "Ensemble learning model based on selected diverse principal components analysis models for process monitoring", *Journal of Chemometrics*, vol.32, no.6, Article No. e3010, 2018. DOI: 10.1002/cem.3010.

[35] M. Sohaib, C. H. Kim, and J. M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis", *Sensors*, vol.17, no.12, Article No. 2876, 2017. DOI: 10.3390/s17122876.

[36] L. Jiang, Z. Song, Z. Ge, and J. Chen, "Robust Self-Supervised Model and Its Application for Fault Detection," *Industrial & Engineering Chemistry Research*, vol. 56, pp. 7503–7515, 2017.

[37] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Deep learning based nonlinear principal component analysis for industrial process fault detection", in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, 2017, pp.1237–1243.

[38] H. Wu, J. Zhao, "Deep convolutional neural network model based chemical process fault diagnosis," *Computers & Chemical Engineering*, vol. 115, pp. 185–197, 2018.

[39] G. E. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems (Ⅰ): Effect of inequality of variance in the one-way classification," *Annals of Mathematical Statistics*, vol. 25, no. 2, pp. 290–302, 1954.

Donglei Zheng received the B.Eng. degree in automation from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2018. He is currently a master student in control science and engineering at Zhejiang University, Hangzhou, China. His major research interest include modeling and fault detection of multi-scale processes.

Le Zhou (M'18) 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, Chung-Yuan Christian University, Chung-Li, Taoyuan, Taiwan, China, from November 2014 to January 2015. 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.

Zhihuan Song received the B.Eng. and M.Eng. degrees in industrial automation from Hefei University of Technology, Anhui, China, in 1983 and 1986, respectively, and the Ph.D. degree in industrial automation from Zhejiang University, Hangzhou, China, in 1997. Since 1997, he has been with the Department of Control Science and Engineering, Zhejiang University, where he was at first a Postdoctoral Research Fellow, then an Associate Professor, and currently a Professor. His research interests include modeling and fault diagnosis of industrial processes, embedded control systems, and advanced process control technologies. He has published more than 200 papers in journals and conference proceedings.