IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 15, 2019, accepted August 23, 2019, date of publication August 27, 2019, date of current version September 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937886

Primary-Auxiliary Statistical Local Kernel Principal Component Analysis and Its Application to Incipient Fault Detection of Nonlinear Industrial Processes

XIAOGANG DENG[1], (Member, IEEE), PEIPEI CAI[1], JIAWEI DENG[1], YUPING CAO[1], AND ZHIHUAN SONG[2]

[1]College of Control Science and Engineering, China University of Petroleum, Qingdao 266580, China
[2]College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

Corresponding author: Xiaogang Deng (dengxiaogang@upc.edu.cn)

This work was supported in part by the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China, under Grant ICT1900306, in part by the Shandong Provincial Key Program of Research and Development under Grant 2018GGX101025, in part by the China Fundamental Research Funds for the Central Universities under Grant 17CX02054, and in part by the National Natural Science Foundation of China under Grant 61403418 and Grant 21606256.

ABSTRACT Statistical local kernel principal component analysis (SLKPCA) has demonstrated its success in incipient fault detection of nonlinear industrial processes by incorporating the statistical local analysis (SLA) technology. However, the basic SLKPCA method builds the statistical model only based on the normal data and neglects the utilization of the prior fault information, which is often available in many industrial cases. To take full advantage of the prior fault information, this paper proposes an enhanced SLKPCA method, called primary-auxiliary SLKPCA (PA-SLKPCA), for better incipient fault monitoring. The contribution of the proposed method includes three aspects. First, one primary-auxiliary statistical monitoring framework is designed, by which not only the normal training data are applied to develop a primary SLKPCA model, but also the prior fault data are used to build the auxiliary SLKPCA models. Second, a double-block modeling strategy is developed to construct the auxiliary SLKPCA model for each fault case, where a variable grouping strategy based on Kullback-Leibler divergence is applied to divide the process variables into the fault-relevant group and fault-independent variable group, and the sub-model is developed for each group. Third, the Bayesian inference is used to combine the statistical results of each variable group, and one weighted fusion strategy is further designed to integrate the monitoring results from the primary and auxiliary models. Lastly, two case studies including one numerical system and the simulated continuous stirred tank reactor (CSTR) system are used for method evaluation and the simulations show that the proposed method can detect the incipient faults effectively and outperform the traditional SLKPCA method.

INDEX TERMS Incipient fault, fault detection, kernel principal component analysis, statistical local analysis, prior fault information.

I. INTRODUCTION

Due to the higher requirements for the safety and continuity in the modern industries, the fault diagnosis technologies become the focus of attention in the field of process system engineering. The advanced supervisory control and data acquisition systems have been widely applied in the modern industrial processes, which make it possible to collect a large amount of running data for process status monitoring. Therefore, many data-driven fault diagnosis methods have been developed [1]–[5], which usually apply the statistical analysis tools, such as principal component analysis (PCA) [6]–[8], canonical correlation analysis [9], and partial least squares [10], to extract the intrinsic features for status description and fault detection. Among these methods, PCA is the most popular one and many enhanced versions have been widely used to deal with the complicated process characteristics including the data nonlinearity, process dynamics,

The associate editor coordinating the review of this article and approving it for publication was Eunil Park.

and multi-mode operation. Jiang and Yan [11] combined the nonlinear kernel PCA (KPCA) with the basic linear PCA to build a parallel PCA-KPCA method, which mines the both linear and nonlinear relationships simultaneously. For the processes with multiple grades, Zhou *et al.* [12] designed a multi-grade PCA (MGPCA) model in a probabilistic framework, and evaluated it by a real industrial polyethylene process. To extract the dynamic information of the components, a dynamic inner PCA (DiPCA) method was presented by Dong and Qin [13] by considering the analysis of time series data. More related studies can be found in some review papers [14]–[16].

The above studies have achieved their success in many applications. However, they may not function well when the incipient fault occurs. Different to the significant/mature fault, which is with the relatively large fault magnitude and leads to the clear process changes, the incipient fault usually means the difficult-to-detect slight abnormalities [17]. Its main characteristics are the small fault magnitude and the underlying damage to process. Because of the small magnitude, the incipient faults are often masked by the process noises. As its influence is unclear and underlying at the initial phase, it is often extremely difficult to detect by the basic PCA and KPCA methods [18]. Based on these characteristics, the incipient fault detection problem is one of the most challenging tasks in the process monitoring field. As time goes on, the incipient fault may evolve into the mature serious fault and lead to the destructive influence. Therefore, timely detection of incipient fault is of great importance to ensure the process safety and improve the product quality.

To handle the incipient fault detection problem, researchers have proposed some modified PCA methods. Considering that the traditional PCA can not detect the small shift effectively without utilizing the process dynamic information, Wold [19] designed an EWMA-PCA method by introducing the exponential weighting moving average strategy. Aiming at the isolation of the incipient faults, Ji *et al.* [20] studied the incipient fault isolation based on the exponential smoothing reconstruction and demonstrated its efficiency by two case studies. To mine the data distribution information for incipient fault detection, Harmouch *et al.* [21] presented a Kullack-Leibler divergence (KLD) based modified PCA method. As to the KLD-PCA method, Chen *et al.* [22] gave some further investigations by defining the symmetrical evaluation function. Chen *et al.* [23] also discussed one probability-relevant PCA method for the incipient detection in the high-speed trains. In order to isolate the incipient fault, Zhao and Gao [24] put forward a sparse dissimilarity analysis method by considering the difference of data distribution. As the incipient fault of batch process is easily concealed by the trend of non-stationary process, a two-layer fault detection model was established by Zhang *et al.* [25], where the cointegration analysis and PCA are combined to monitor the nonstationary and stationary variables, respectively. Ge *et al.* [26] has put forward a two-step fault detection method considering both significant faults and incipient faults, which adopts traditional multivariate statistical monitoring method to detect significant faults, and proposes an optimal residual generation method to detect incipient faults. For detecting the incipient sensor faults in the multimode processes, a modified recursive transformed component statistical analysis method was developed by Shang *et al.* [27] by utilizing conditionally independent Bayesian learning.

Besides these above-mentioned methods, an important tool for incipient fault monitoring is the statistical local analysis (SLA) [28]. SLA transforms the complex fault detection into the monitoring of the mean of a Gaussian vector and constructs an improved residual for inspecting the intrinsic data changes. Kruger *et al.* [29] firstly combined it with PCA to provide the statistical local PCA (SLPCA) for detecting the early changes of the covariance structure of the process variables. It has been proved that SLPCA can enhance the ability of traditional PCA to monitor incipient fault. Furthermore, Chen *et al.* [30] and Li *et al.* [31] also validated that SLA can be propitious to the sensitivity of incipient fault detection. To handle the issue of the nonlinear process incipient fault monitoring, Ge *et al.* [32] developed the statistical local kernel PCA (SLKPCA) by integrating SLA with KPCA and the applications show that SLKPCA is more effective than the traditional KPCA in terms of incipient fault detection of the benchmark Tennessee Eastman process. Deng and Deng [33] and Wang [34] further discussed the combination of SLA with kernel methods.

Considering that the advantages of SLKPCA, this paper is to make SLKPCA as a basic tool and seek for a modified SLKPCA method for better incipient fault monitoring performance. One clear shortcoming of the present SLKPCA methods is the neglecting of the available prior fault information. In the usual SLKPCA modeling procedure, only the normal data are used for statistical analysis with the assumption of no available prior fault information. In fact, there are often some available fault information in the real industrial cases. The fault information includes the prior fault data, the fault mechanism and the expert knowledge, etc. Especially, the prior fault data can be very helpful for fault detection. Although the fault data may be not enough to cover all the possible patterns, they still contain many important information which is very helpful to detect the incipient faults, especifically the incipient faults same to the known fault pattern in the database. Therefore, to build an auxiliary model by the prior fault data is of great value to improve the incipient fault detection.

Based on these discussions, this paper is to propose an enhanced SLKPCA method, referred to as primary-auxiliary SLKPCA model, for better incipient fault detection performance by making good use of the prior fault data. In the proposed method, the traditional SLKPCA model based on the normal data is used as the primary monitoring model to ensure the performance baseline. By analyzing the fault data, an auxiliary SLKPCA model is developed which applies the Kullback-Leibler divergence to divide the variables into the fault-related group and fault-independent group, and builds a

sub-model for each group. Furthermore, the Bayesian inference technology is applied to combine the monitoring results from each auxiliary sub-models, and the weighted fusion strategy is designed to integrate the results from the primary and auxiliary SLKPCA models. Finally, two case studies are used to show the effectiveness of the proposed method.

The rest of this paper is summarized as follows. In section II, traditional KPCA and SLKPCA are reviewed briefly. Section III gives a specific variable grouping strategy based on KLD and details the principle of the primary-auxiliary SLKPCA (PA-SLKPCA) method. Section IV describes the fault detection framework and the corresponding flow chart. Then, two simulation examples are presented to illustrate the proposed method in section V. Finally, some conclusions are drawn in section VI.

II. PRELIMINARIES

A. KPCA

As a nonlinear extension of the linear PCA method, KPCA is an effective nonlinear process monitoring tool [35]–[37]. The basic idea of this method is to transform the original variables into high-dimensional feature space by an implicit nonlinear mapping function, and then implement the standard linear PCA in the feature space [35], [38]. Compared with other nonlinear PCA methods, its advantage lies in avoiding the complex nonlinear optimization problem by introducing the kernel trick. The fundamental principle of this method is described as follows.

For the given training data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$, where n and m are the sample number and the variable number, respectively, a nonlinear function $\Psi(\cdot)$ is used to map the data $\boldsymbol{X}$ onto the new feature space $\Psi(\boldsymbol{X})$, where the variables are linearly related. In the linear feature space, the PCA decomposition is carried out as

$$\Psi(\boldsymbol{X}) = \sum_{j=1}^{k} \boldsymbol{t}_j \boldsymbol{v}_j^T + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{v}_j$ is called the loading vector, $\boldsymbol{t}_j$ is called the score vector or principal component vector, $\boldsymbol{E}$ is the residual matrix, and k is the the retained principal component number.

The loading vector $\boldsymbol{v}_j$ represents the data projection direction, which can be computed by the eigenvalue decomposition of the $\Psi(\boldsymbol{X})$'s covariance matrix as

$$\frac{1}{n-1}\Psi^T(\boldsymbol{X})\Psi(\boldsymbol{X})\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j, \tag{2}$$

where λ_j is the j-th eigenvalue corresponding to the eigenvector $\boldsymbol{v}_j$. As the nonlinear mapping function $\Psi(\cdot)$ is usually unknown, it is difficult to solve Eq. (2). Considering that the projection vector $\boldsymbol{v}_j$ can be expanded by the training data matrix $\Psi(\boldsymbol{X})$ as [38]

$$\boldsymbol{v}_j = \Psi^T(\boldsymbol{X})\boldsymbol{p}_j, \tag{3}$$

where $\boldsymbol{p}_j \in \mathbb{R}^n$ is the coefficient vector.

By combining Eqs.(2) and (3), the KPCA optimization will be lastly transformed into an kernel matrix decomposition as

$$\frac{1}{n-1}\boldsymbol{K}\boldsymbol{p}_j = \lambda_j \boldsymbol{p}_j, \tag{4}$$

where $\boldsymbol{K}$ is the kernel matrix defined by $\boldsymbol{K} = \Psi(\boldsymbol{X})\Psi^T(\boldsymbol{X})$, whose element k_{ij} is the inner product of the i-th sample $\Psi(\boldsymbol{x}_i)$ and the j-th sample $\Psi(\boldsymbol{x}_j)$. That means $k_{ij} = \Psi^T(\boldsymbol{x}_i)\Psi(\boldsymbol{x}_j)$. To compute the k_{ij}, the kernel trick is applied as

$$k_{ij} = \Psi^T(\boldsymbol{x}_i)\Psi(\boldsymbol{x}_j) = ker(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{5}$$

In this paper, we adopt the commonly used radius basis kernel function $ker(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-||\boldsymbol{x}_i - \boldsymbol{x}_j||^2/\sigma)$, where σ is the kernel width. By solving Eq.(5), a series of eigenvectors $\boldsymbol{p}_j$ ordered by their corresponding eigenvalues λ_j are available. Usually, we only retain the first $\overline{n}$ non-zero eigenvalues for statistical modeling. In this paper, $\overline{n}$ is selected so that the cumulative eigenvalue sum exceeds 99.9% of the overall cumulative eigenvalue sum.

For a testing data sample $\boldsymbol{x}_h \in \mathbb{R}^m$ at the $h-$th sample instant, its j-th kernel principal component $t_{h,j}$ can be obtained by projecting the $\Psi(\boldsymbol{x}_h)$ onto the eigenvector $\boldsymbol{v}_j$, which is expressed by

$$t_{h,j} = \Psi^T(\boldsymbol{x}_h)\boldsymbol{v}_j = \boldsymbol{K}_h^T \boldsymbol{p}_j, \tag{6}$$

where $\boldsymbol{K}_h = \Psi(\boldsymbol{X})\Psi(\boldsymbol{x}_h) \in \mathbb{R}^n$ is the kernel vector of the testing vector $\boldsymbol{x}_h$.

Two monitoring statistics T^2 and Q are used to monitor the process changes. T^2 statistic represents the variation in the kernel principal component subspace, depicted by the first k kernel principal components, while Q statistic monitors the data changes in the residual subspace, represented by the other kernel principal components. For the testing vector $\boldsymbol{x}_h$, the corresponding statistics are developed as

$$T^2 = [t_{h,1},\ t_{h,2},\ \cdots,\ t_{h,k}]\Lambda_k^{-1}[t_{h,1},\ t_{h,2},\ \cdots,\ t_{h,k}]^T, \tag{7}$$

$$Q = \sum_{j=1}^{\overline{n}} t_{h,j}^2 - \sum_{j=1}^{k} t_{h,j}^2, \tag{8}$$

where Λ_k is the diagonal matrix composed of the first k eigenvalues $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k$. The confidence limits of the T^2 and Q statistics are determined by the non-parametric kernel density estimation (KDE) technology [39]. When one of these two statistics exceeds the corresponding confidence limit, it means one fault is detected.

B. STATISTICAL LOCAL KPCA

Statistical local analysis (SLA) has proven to be an effective approach to detect the abrupt data changes. SLA investigates the changes of model parameters of a given system function. The parameter deviation can be detected by a change in the mean value of a Gaussian probability density function [28], [29], [32]. By combining SLA with KPCA, an enhance KPCA method – Statistical local KPCA (SLKPCA) can be developed [32]. Based on the scores vectors and the eigenvalues, SLKPCA constructs a novel residual function, which has

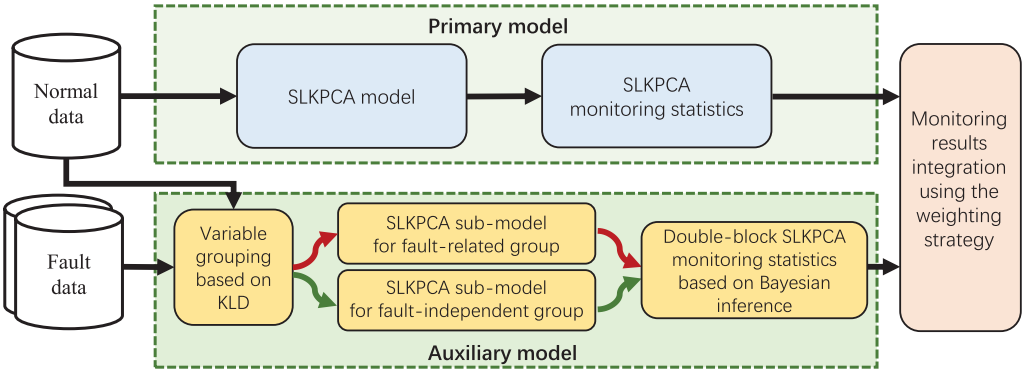


FIGURE 1. The modeling framework of the proposed PA-SLKPCA method.

a more powerful capability in the detection of incipient faults. For the kernel principal component $t_{h,j}$, the corresponding initial residual is defined as [32]:

$$r_{h,j} = 2t_{h,j}^2 - 2\lambda_j \tag{9}$$

Next, the improved residual is built by applying the moving window technology, which is expressed as [32]

$$\widetilde{r}_{h,j} = \frac{1}{\sqrt{w}} \sum_{l=h-w+1}^{h} (2t_{l,j}^2 - 2\lambda_j) \tag{10}$$

where w is the width of the moving window. w is an important parameter for SLKPCA. A large w helps increasing the detection rate of incipient fault because of the introduction of the historical data. However too large w may lead to the large detection time delay. The reasonable value of w can be selected by experiment testing.

For one testing vector $\boldsymbol{x}_h$, its corresponding improved residual vector is denoted by $\widetilde{\boldsymbol{r}}_h = [\widetilde{r}_{h,1}\ \widetilde{r}_{h,2}\ \cdots\ \widetilde{r}_{h,\bar{n}}]$, where the first k components describe the principal system changes, while the other $\bar{n}-k$ components represent the noise information. Therefore, the improved residual vector $\tilde{\boldsymbol{r}}_h$ is divided into two parts $\widetilde{\boldsymbol{rs}}_h$ and $\widetilde{\boldsymbol{rn}}_h$ expressed by

$$\widetilde{\boldsymbol{rs}}_h = [\widetilde{r}_{h,1}\ \widetilde{r}_{h,2}\ \cdots\ \widetilde{r}_{h,k}] \tag{11}$$

$$\widetilde{\boldsymbol{rn}}_h = [\widetilde{r}_{h,k+1}\ \widetilde{r}_{h,k+2}\ \cdots\ \widetilde{r}_{h,\bar{n}}] \tag{12}$$

Based on the improved residuals $\widetilde{\boldsymbol{rs}}_h$ and $\widetilde{\boldsymbol{rn}}_h$, two monitoring statistics of SLKPCA are developed as

$$T^2 = \widetilde{\boldsymbol{rs}}_h \Sigma_s^{-1} (\widetilde{\boldsymbol{rs}}_h)^T, \tag{13}$$

$$Q = \widetilde{\boldsymbol{rn}}_h \Sigma_n^{-1} (\widetilde{\boldsymbol{rn}}_h)^T, \tag{14}$$

where Σ_s and Σ_n are the residual covariance matrix, which can be computed by the training data.

As an unsupervised fault detection method, SLKPCA improves the performance of traditional KPCA in monitoring incipient faults. However, this method only uses normal data in statistical modeling without considering the utilization of prior fault information, which will result in the omission and waste of some prior fault information, and further lead to the degradation of fault detection performance. In this paper, we try to make good use of the prior fault data to improve the basic SLKPCA method.

III. PRIMARY-AUXILIARY STATISTICAL LOCAL KPCA

In order to make full use of the prior fault data, we are to design a modified SLKPCA approach, called primary-auxiliary SLKPCA (PA-SLKPCA), which combines the normal training data and the prior fault data simultaneously for statistical modeling. The whole modeling framework is shown in Fig. 1. The normal training data are applied to build the primary SLKPCA model, which plays a dominating role in the process monitoring. Besides, the prior fault data are used to develop the auxiliary SLKPCA models, which focus on the emphasis of the fault information and assist the monitoring of incipient faults. The monitoring results from the primary model and the auxiliary models are integrated based on the weighting strategy for a comprehensive monitoring. Different to the traditional SLKPCA methods, the proposed PA-SLKPCA method shows the great potential in the detection of the incipient faults because of the introduction of the prior fault information.

As the SLKPCA modeling has been discussed in the previous section, this section focuses on the details of the auxiliary SLKPCA modeling. The auxiliary SLKPCA model is in fact a double-block model, which firstly divides all the variables into two blocks including the fault-related group and the fault-independent group, and then builds the sub-model for each group, respectively. Specifically, the auxiliary SLKPCA modeling includes three steps of variable grouping, sub-model development and monitoring statistics integration. We describe the details in the following.

A. VARIABLE GROUPING USING KL DIVERGENCE

The first step of the auxiliary SLKPCA modeling is to distinguish the fault-related variables from the other variables, so that the developed auxiliary model targets the specific fault and provides more pertinent information for incipient fault detection. To handle this problem, we propose a novel variable grouping strategy based on KL divergence, which

measures the difference of probability distributions of two variables to locate the fault variables.

Among all the monitored variables, some variables involving the fault are called fault-related variables, while other variables are not related or weakly related to the fault, called fault-independent variables. Usually, it is reasonable to suppose that the fault-related variables have the different probability distributions compared to the normal variables, while the fault-independent variables are very similar to the normal one in terms of the probability distribution.

Kullback-Leibler divergence (KL divergence), also called relative entropy or information divergence, is an effective tool to evaluate the difference between two probability density functions [40]–[42]. Given two continuous probability density functions $f(x)$ and $g(x)$ on a random variable x, the KL information is defined as [40]

$$I(f||g) = \int f(x) log \frac{f(x)}{g(x)} dx. \tag{15}$$

Although the KL information seems to be a distance measure between two distributions, it is not a real distance because of its asymmetric property. Furtherly, the KL divergence is built as the symmetric form of the KL information, which is expressed by

$$KLD(f, g) = I(f||g) + I(g||f). \tag{16}$$

If $f(x)$ and $g(x)$ are from the same distributions, the KL divergence is equal to zero. Otherwise, when these two distributions are different, the KL divergence will be a value bigger than zero. Greater the KL divergence is, more different the two distributions are. By assuming the Gaussian distribution for $f(x)$ and $g(x)$, their probability density functions are given as

$$f(x) = \frac{1}{\sqrt{2\pi\lambda_f}} exp(-\frac{(x-\mu_f)^2}{2\lambda_f}), \tag{17}$$

$$g(x) = \frac{1}{\sqrt{2\pi\lambda_g}} exp(-\frac{(x-\mu_g)^2}{2\lambda_g}), \tag{18}$$

where μ_f, λ_f are the mean and variance of $f(x)$, respectively, while μ_g, λ_g are the mean and variance of $g(x)$, respectively.

With Eqs. (17) and (18), the KL information can be computed as [22]

$$I(f||g) = \frac{1}{2}[log(\frac{\lambda_g}{\lambda_f}) + \frac{\lambda_f}{\lambda_g} + \frac{(\mu_f - \mu_g)^2}{\lambda_g} - 1], \tag{19}$$

and then the KL divergence is given by [22]

$$KLD(f, g) = \frac{1}{2}[\frac{\lambda_f}{\lambda_g} + \frac{\lambda_g}{\lambda_f} + (\mu_f - \mu_g)^2(\frac{1}{\lambda_f} + \frac{1}{\lambda_g}) - 2]. \tag{20}$$

Given one prior fault dataset $\boldsymbol{X}^f \in \boldsymbol{R}^{n_f \times m}$ and the normal dataset $\boldsymbol{X} \in \boldsymbol{R}^{n \times m}$, they are with the same m measured variables $x_1, x_2, \cdots, x_m$. Assume that the data distribution $f(x_i)$ describes the normal data while the data distribution $g(x_i)$ is for the fault data, then the KL divergence of the variable x_i is computed by

$$KLD_i = \frac{1}{2}[\frac{\lambda_f(x_i)}{\lambda_g(x_i)} + \frac{\lambda_g(x_i)}{\lambda_f(x_i)} + (\mu_f(x_i) - \mu_g(x_i))^2 (\frac{1}{\lambda_f(x_i)} + \frac{1}{\lambda_g(x_i)}) - 2]. \tag{21}$$

If KLD_i is clearly bigger than zero, that means these two distributions are very different and the variable x_i should be regarded as a fault-related variable. Otherwise, if the KL divergence is very close to zero, the corresponding variable is a fault-independent variable. In the real applications, the random noises affect the computation of KL divergence so that the KL divergence between two same data distributions are not exactly zero. Therefore, we give the following rules for variable grouping

$$\begin{cases} x_i \, is \, fault - related, & if \; KLD_i \geq \sigma_i, \\ x_i \, is \, fault - indepdent, & if \; KLD_i < \sigma_i, \end{cases} \tag{22}$$

where σ_i is the KL divergence threshold for variable grouping. The determination of σ_i is data-driven. Firstly the normal KLD_i values corresponding to the variable x_i are computed and then the KDE technique is applied to calculate the 99% confidence limit as the threshold σ_i.

By this decision rule, the training data matrix $\boldsymbol{X}$ can be divided into two blocks $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, which are corresponding to the fault-related variable group and the fault-independent variable group, respectively.

It should be noted that an implicit assumption is the probability distribution difference can be observed in the given prior fault. In this paper, we only consider the case of the significant prior faults, which meet the above assumption.

B. AUXILIARY MODEL DEVELOPMENT

For each prior fault pattern, the monitored variables can be divided into two groups and then the SLKPCA modeling is performed for each group. So a double-block SLKPCA model is obtained based on one prior fault dataset. If there are C prior fault patterns, the variable grouping can be carried out C times, which leads to C auxiliary double-block SLKPCA models. For the c-th fault case, its double-block SLKPCA modeling brings the improved residuals expressed as

$$\widetilde{r}_{h,j}^{(c,b)} = \frac{1}{\sqrt{w}} \sum_{l=h-w+1}^{h} 2(t_{h,j}^{(c,b)})^2 - 2\lambda_j^{(c,b)}, \tag{23}$$

where $c = 1, 2, \cdots, C$ indicates the fault case number, while $b = 1, 2$ is the index of sub-model in the double-block SLKCPA model. $b = 1$ corresponds to the fault-related variable group, while $b = 2$ indicates the fault-independent variable group.

All the residual components are divided into two parts of $\widetilde{\boldsymbol{rs}}_h$ for the principal system change part and $\widetilde{\boldsymbol{rn}}_h$ for the noise

part, written by

$$\widetilde{\boldsymbol{rs}}_h^{(c,b)} = [\widetilde{r}_{h,1}^{(c,b)}\ \widetilde{r}_{h,2}^{(c,b)}\ \cdots\ \widetilde{r}_{h,k}^{(c,b)}], \tag{24}$$

$$\widetilde{\boldsymbol{rn}}_h^{(c,b)} = [\widetilde{r}_{h,k+1}^{(c,b)}\ \widetilde{r}_{h,k+2}^{(c,b)}\ \cdots\ \widetilde{r}_{h,\bar{n}}^{(c,b)}]. \tag{25}$$

Based on the residual vectors, the SLKPCA sub-model monitoring statistics T^2 and Q are constructed as

$$T^{2(c,b)} = \widetilde{\boldsymbol{rs}}_h^{(c,b)}(\Sigma_s^{(c,b)})^{-1}(\widetilde{\boldsymbol{rs}}_h^{(c,b)})^T, \tag{26}$$

$$Q^{(c,b)} = \widetilde{\boldsymbol{rn}}_h^{(c,b)}(\Sigma_n^{(c,b)})^{-1}(\widetilde{\boldsymbol{rn}}_h^{(c,b)})^T, \tag{27}$$

where $T_h^{2(c,b)}$ and $Q_h^{(c,b)}$ correspond to the b-th sub-model for the double-block SLKPCA model of the c-th fault case. For the C prior fault datasets, there will be C auxiliary SLKPCA models. All these auxiliary models are integrated together to become one holistic auxiliary model. The integration strategy is discussed in the following section.

C. MONITORING RESULTS INTEGRATION

After obtaining the monitoring statistics of each group, Bayesian fusion strategy is applied to integrate the monitoring results of the sub-models.

For one testing sample $\boldsymbol{x}_h$, its fault probabilities, decided by the T^2 statistic from the b-th sub-model of the c-th double-block SLKPCA model, are denoted by $p_{T^2}^{(c,b)}(F|\boldsymbol{x}_h)$, which can be computed by the Bayesian inference as [43], [44]

$$p_{T^2}^{(c,b)}(F|\boldsymbol{x}_h) = \frac{p_{T^2}^{(c,b)}(\boldsymbol{x}_h|F)p(F)}{p_{T^2}^{(c,b)}(\boldsymbol{x}_h|N)p(N) + p_{T^2}^{(c,b)}(\boldsymbol{x}_h|F)p(F)}, \tag{28}$$

where $p(F)$ and $p(N)$ are the prior fault and normal probabilities, respectively. $p(F)$ is equivalent to the significance level α while $p(N)$ is equal to the confidence level $1-\alpha$, respectively. $p_{T^2}^{(c,b)}(\boldsymbol{x}_h|F)$ and $p_{T^2}^{(c,b)}(\boldsymbol{x}_h|N)$ are the posterior probabilities under fault and normal conditions, respectively, which can be calculated as

$$p_{T^2}^{(c,b)}(\boldsymbol{x}_h|N) = exp\left(-\frac{T^{2(c,b)}}{T_{lim}^{2\ (c,b)}}\right) \tag{29}$$

$$p_{T^2}^{(c,b)}(\boldsymbol{x}_h|F) = exp\left(-\frac{T_{lim}^{2\ (c,b)}}{T^{2(c,b)}}\right) \tag{30}$$

where $T_{lim}^{2\ (c,b)}$ is the confidence limit of the monitoring statistic $T_h^{2(c,b)}$.

Similarly, the fault probabilities for the Q statistic is expressed by [43], [44]

$$p_Q^{(c,b)}(F|\boldsymbol{x}_h) = \frac{p_Q^{(c,b)}(\boldsymbol{x}_h|F)p(F)}{p_Q^{(c,b)}(\boldsymbol{x}_h|N)p(N) + p_Q^{(c,b)}(\boldsymbol{x}_h|F)p(F)} \tag{31}$$

where the posterior probabilities $p_Q^{(c,b)}(\boldsymbol{x}_h|F)$ and $p_Q^{(c,b)}(\boldsymbol{x}_h|N)$ are obtained by

$$p_Q^{(c,b)}(\boldsymbol{x}_h|N) = exp\left(-\frac{Q^{(c,b)}}{Q_{lim}{}^{(c,b)}}\right) \tag{32}$$

$$p_Q^{(c,b)}(\boldsymbol{x}_h|F) = exp\left(-\frac{Q_{lim}{}^{(c,b)}}{Q^{(c,b)}}\right) \tag{33}$$

By combining the monitoring statistics from the fault-related group and the fault-independent group, two unified fault probability indices for the c-th auxiliary model are developed as

$$BIC_{T^2}^{(c)} = \sum_{b=1}^{2} p_{T^2}^{(c,b)}(F|\boldsymbol{x}_h)\frac{p_{T^2}^{(c,b)}(\boldsymbol{x}_h|F)}{\sum_{l=1}^{2} p_{T^2}^{(c,l)}(\boldsymbol{x}_h|F)} \tag{34}$$

$$BIC_Q^{(c)} = \sum_{b=1}^{2} p_Q^{(c,b)}(F|\boldsymbol{x}_h)\frac{p_Q^{(c,b)}(\boldsymbol{x}_h|F)}{\sum_{l=1}^{2} p_Q^{(c,l)}(\boldsymbol{x}_h|F)} \tag{35}$$

So far, both the primary and the auxiliary SLKPCA models are built with their statistics as T^2, Q, $BIC_{T^2}^{(c)}$ and $BIC_Q^{(c)}$. The following question is how to combine the results from the primary-auxiliary models for the holistic monitoring. A natural idea is to apply one reasonable weighting strategy. As the primary SLKPCA model is the basic model, it should always works. Because the auxiliary model only aims at some specific known fault, it performs well when the corresponding fault occurs and can not ensure a good performance for other kinds of faults. Therefore, the auxiliary model for one specific fault should be activated only when it detects the fault. Based on these considerations, we design a weighting approach as follows.

$$WT^2 = \frac{T^2}{T_{lim}^2} + \sum_{c=1}^{C} w_{T^2}^{(c)}\frac{BIC_{T^2}^{(c)}}{\alpha} \tag{36}$$

$$WQ = \frac{Q}{Q_{lim}} + \sum_{c=1}^{C} w_Q^{(c)}\frac{BIC_Q^{(c)}}{\alpha} \tag{37}$$

where $w_{T^2}^{(c)}$ is set 1 if the $BIC_{T^2}^{(c)}$ statistics of several continuous samples are beyond the significance level α, otherwise it is zero. $w_Q^{(c)}$ is set by the same rule. The weighted statistics WT^2 and WQ can give a comprehensive and sensitive detection for the incipient faults. When $WT^2 > 1$ or $WQ > 1$, the process is considered to be faulty. Otherwise, the process is under normal condition.

IV. THE PROCESS MONITORING PROCEDURE USING PA-SLKPCA

The incipient fault detection method based on PA-SLKPCA can be divided into two stages: offline modeling and online monitoring, and the corresponding flow chart is shown in Fig. 2. The detailed steps are described below.

offline modeling

(1) Collect the normal training data $\boldsymbol{X}$ and the C prior fault datasets $\boldsymbol{X}^c(c = 1, 2, \cdots, C)$ from the historical datasets, and standardize them using the mean and variance of the normal data.

(2) Establish the primary SLKPCA model using the normal data $\boldsymbol{X}$.

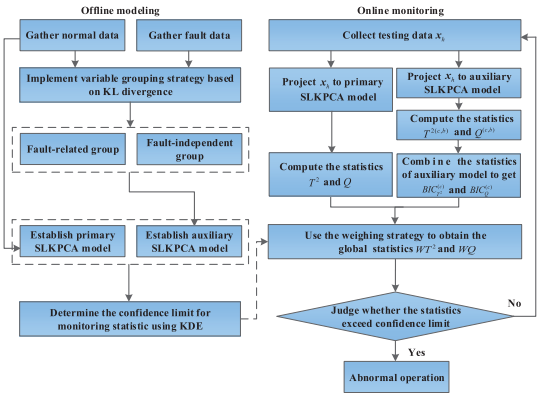


FIGURE 2. The process monitoring flowchart based on PA-SLKPCA.

(3) For each fault dataset $\boldsymbol{X}^c$, perform the variable grouping and build an double-block SLKPCA model as the c-th auxiliary model.
(4) For some normal validation data, compute the monitoring statistics of the primary and the auxiliary SLKPCA models, respectively, and determine the monitoring statistics' confidence limits T^2_{lim}, Q_{lim}, $T^{2\ (c,b)}_{lim}$ and $Q_{lim}{}^{(c,b)}$ using the KDE technology.

Online monitoring

(1) Gather the testing data $\boldsymbol{x}_h$ at the h-th sample instant, and scale it with the mean and variance of the normal data.
(2) Project the normalized data on the primary SLKPCA model and all the auxiliary SLKPCA models, and compute the monitoring statistics T^2, Q, $T^{2(c,b)}$ and $Q^{(c,b)}$.
(3) Compute the integrated statistics $BIC^{(c)}_{T^2}$ and $BIC^{(c)}_{Q}$ for each auxiliary model using the Bayesian inference strategy.
(4) Obtain the final primary-auxiliary model statistics WT^2 and WQ and determine if some fault occurs.

V. CASE STUDIES

In this section, a numerical example and the simulated continuous stirred tank reactor (CSTR) system are used to verify the proposed algorithm. The traditional KPCA and SLKPCA are also applied for comparison. In order to evaluate the different methods, three indices including fault detection rate (FDR), fault detection time (FDT) and false alarm rate (FAR) are adopted. FDR is the percentage of faulty samples exceeding the confidence limit over the total faulty samples, FDT refers to the first sample of six consecutive faulty samples exceeding the confidence limit, while FAR is the ratio of the number of the normal samples over the confidence limit to the total normal sample number. In the monitoring charts of all methods, the confidence limits are plotted in the dashed line while the statistics are given by the solid line. For the convenience of method comparison, all the confidence limits in the monitoring charts are normalized to 1.

A. A NUMERICAL EXAMPLE

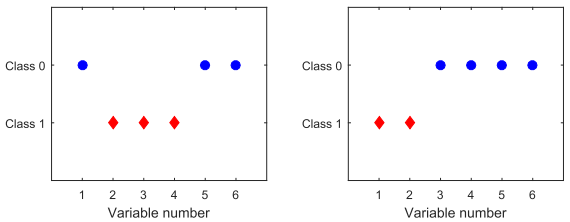
In this paper, a nonlinear numerical system with 6 variables is designed with the following mathematical equations.

$$\begin{cases} x_1 = t_1 + e_1, \\ x_2 = t_1^2 - 2t_2 + e_2, \\ x_3 = -t_2^3 + 3t_3^2 + e_3, \\ x_4 = t_2 + 5t_4 + e_4, \\ x_5 = t_3^2 - 2t_4 + e_5, \\ x_6 = -t_3^2 + 3t_4^3 + e_6, \end{cases} \tag{38}$$

where the data source $t_i (1 \leq i \leq 4) \in [0, 2]$ is the random signal with uniform distribution, while $e_i (1 \leq i \leq 6)$ is the Gaussian-distributed noise with the zero mean and the variance of 0.3. First, 300 samples under the normal condition are generated for model training while another 2000 normal samples are as the validation dataset for the confidence limit determination. Then two groups of incipient faults including one small step change and one weak ramp change, tabulated as follows, are simulated for testing the performance of incipient fault detection methods. Each fault dataset includes 500 samples.

- Fault D1: t_2 is imposed with a step bias of 0.5 from the 201st sample.
- Fault D2: t_1 is injected with a ramp change with the slope of 0.005 from the 201st sample.

Three methods of KPCA, SLKPCA, PA-SLKPCA are applied for statistical modeling. For all methods, the kernel width σ is set to 100 and the retained KPC number is computed by the 90% cumulative percent variance (CPV) criterion. For the SLKPCA and PA-SLKPCA methods, the moving window width w is set to 20. The 99% confidence limits, obtained by the KDE method, are adopted for all statistics. That means the significance level α is equal to 0.01. When PA-SLKPCA is applied, two prior fault datasets are generated. It should be noted that the prior fault are supposed to be the significant type in this paper. The step bias of the prior fault D1 is set to 2.0, while the slop value of the prior fault D2 is set to 0.02. The variable group results for the fault D1 and D2 are demonstrated in the Fig. 3, where Class 0 indicates the fault-independent variables, while Class 1 represents the fault-related variables. From the Fig. 3(a), it can be observed that the fault-related variables correspond to the Nos. 2, 3 and 4, which are the same to the real faulty variables.



(a) Fault D1 (b) Fault D2

FIGURE 3. Variable grouping results for the fault D1 and D2.

By the Fig. 3(b), the fault-related variables are indicated as the Nos. 1, and 2. This also corresponds to the real fault mechanism. By applying these prior fault information, the auxiliary SLKPCA monitoring model can be built.

Firstly, the monitoring charts on the fault D1 using KPCA, SLKPCA and PA-SLKPCA are illustrated in Figs. 4, 5, and 6. In this case, when PA-SLKPCA is applied, only the prior information of fault D1 is utilized. The corresponding monitoring indices are listed in the Table 1, where – represents fault can not be detected by this statistic. The KPCA monitoring charts in the Fig. 4 indicate that the fault D1 can not be detected effectively. It has the poor FDRs of 0.08 and 0.12 for the T^2 and Q statistics, respectively. When SLKPCA is applied, it improves the fault detection performance clearly. The FDRs of the T^2 and Q statistics are increased to 0.493 and 0.767, respectively. By applying the PA-SLKPCA with the prior information of fault D1, the WT^2 FDR gets to the 0.663, while the WQ FDR is also enhanced to the 0.853. By the Table 1, three methods have the similar FARs, while the FDTs of SLKPCA and PA-SLKPCA are superior to the basic KPCA method. To further analyze the PA-SLKPCA method, the monitoring charts in the auxiliary SLKPCA (A-SLKPCA) model are displayed. Fig. 7(a) gives the SLKPCA monitoring charts from the fault-related sub-model, while Fig. 7(b) demonstrates the monitoring results

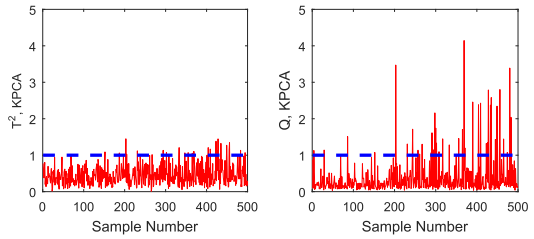


FIGURE 4. The KPCA monitoring charts on the fault D1.

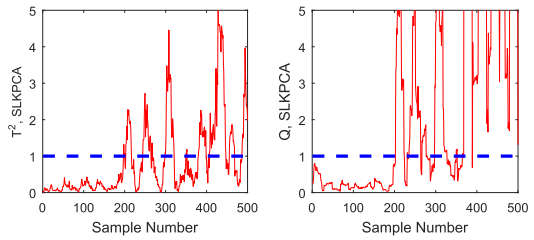


FIGURE 5. The SLKPCA monitoring charts on the fault D1.

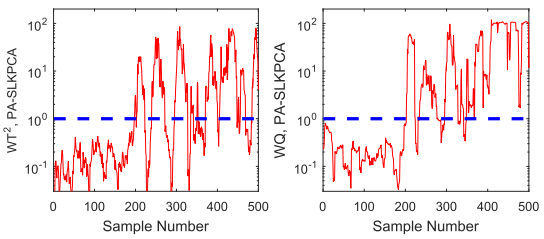
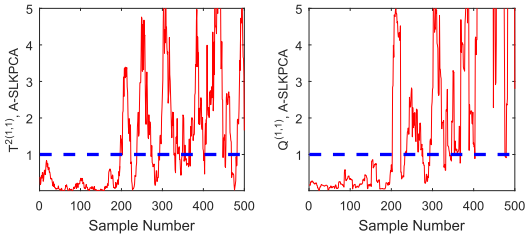


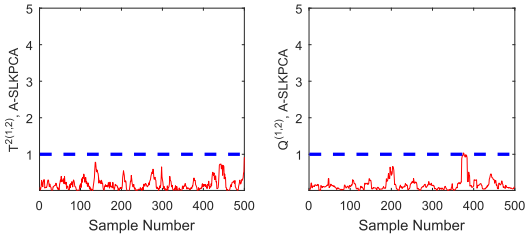
FIGURE 6. The PA-SLKPCA monitoring charts on the fault D1.

TABLE 1. Monitoring results on fault D1 dataset obtained by KPCA, SLKPCA and PA-SLKPCA.

Method	KPCA		SLKPCA		PA-SLKPCA	
Statistic	T^2	Q	T^2	Q	WT^2	WQ
FDR	0.08	0.12	0.493	0.767	0.663	0.853
FDT	–	–	203	203	203	203
FAR	0.01	0.02	0.01	0	0.015	0



(a) fault-related group



(b) fault-independent group

FIGURE 7. The auxiliary SLKPCA monitoring charts on the fault D1.

from the fault-independent sub-model. The T^2 and Q's FDRs of fault-related sub-model are 0.737 and 0.843, respectively, which are much higher than the SLKPCA's. By contrast, Fig. 7(b) gives no clear fault alarms. That shows the auxiliary SLKPCA model based on the double-block modeling strategy can emphasize the influence of fault variables and drives the performance increasing. All these analyses demonstrate that the utilization of the prior fault information benefits the detection of the incipient faults significantly.

The former PA-SLKPCA method only considers the prior information of fault D1. Does it help or worsen the detection of the fault D2? The conclusion can be observed from the Figs. 8, 9, and 10, and Table 2. The KPCA detection rates on the fault D2 are 0.197 and 0.143 for T^2 and Q, respectively, which shows the unsatisfactory performance. In the Fig. 9, SLKPCA does better, which prompts the indices to 0.473 and 0.656, respectively. When PA-SLKPCA is applied, the WT^2 and WQ detection rates are 0.493 and 0.680, respectively, which are not the remarkable improvement. This is because the prior information of fault D2 is not considered in the PA-SLKPCA. However, it should be pointed out that the monitoring performance is not worsened even if the corresponding fault information is not considered. Furthermore, we add the prior information of fault D2 in the auxiliary model. The augmented method is denoted as PA-SLKPCA-2. Compared to the PA-SLKPCA, the PA-SLKPCA2 gives a little enhancement on the WT^2 statistics with the FDR of 0.513 while

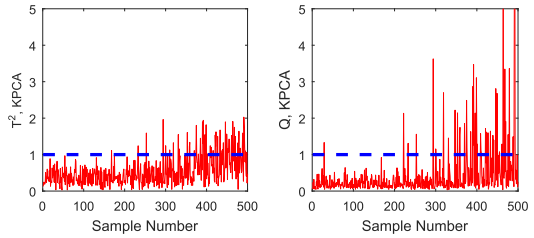


FIGURE 8. The KPCA monitoring charts on the fault D2.

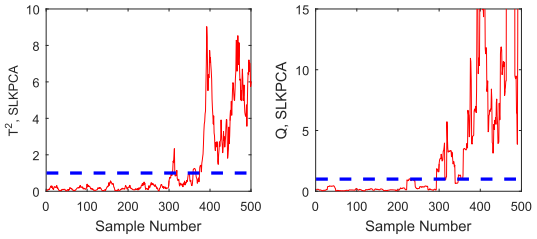


FIGURE 9. The SLKPCA monitoring charts on the fault D2.

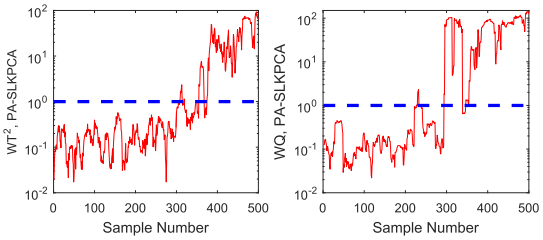


FIGURE 10. The monitoring charts on the fault D2 based on PA-SLKPCA.

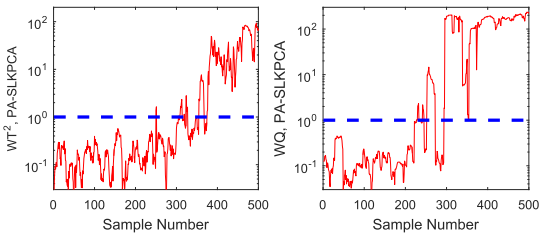


FIGURE 11. The monitoring charts of the fault D2 based on PA-SLKPCA-2.

TABLE 2. Monitoring results on fault D2 dataset obtained by KPCA, SLKPCA and PA-SLKPCA.

Method	KPCA		SLKPCA		PA-SLKPCA		PA-SLKPCA-2	
Statistic	T^2	Q	T^2	Q	T^2	Q	WT^2	WQ
FDR	0.197	0.143	0.473	0.656	0.493	0.680	0.513	0.790
FDT	–	–	308	294	308	294	308	252
FAR	0.005	0.005	0.	0	0	0	0	0

shows a clear improvement on the WQ statistic with the FDR of 0.790. Also, the PA-SLKPCA-2 WQ statistic detects the fault at the 252nd sample, while the SLKPCA Q statistic only detects the fault at the 294th sample. This means that PA-SLKPCA modeling can improve the SLKPCA method if the corresponding fault information is utilized.

To sum up, PA-SLKPCA can detect the incipient fault D1 and D2 more effectively with the utilization of the prior fault information. More prior fault information is available, better monitoring performance are obtained. When the fault information is not available, PA-SLKPCA still works and its performance is the same to the basic SLKPCA method.

B. THE CSTR SYSTEM

Continuous stirred tank reactor (CSTR) is a commonly used equipment in the field of chemical industry, which has been extensively applied as a benchmark system to validate the fault detection and diagnosis methods [45]–[47]. The schematic diagram of one CSTR system is shown in Fig. 12, where the first-order irreversible exothermic reaction $A \rightarrow B$ occurs and the cooling jacket removes the reaction heat. To make the reaction temperature stable, one cascade control system is designed which applies the reaction temperature controller as the master controller and the coolant flow controller as the slave controller. Also a level cascade control system is designed by using the output flow controller as the slave controller. In the Fig 12, Q_C denotes the coolant flow, Q_F denotes the reactor feed flow, Q_O denotes the reactor output flow, T_F represents the reactor feed temperature, T represents the reactor temperature, T_{CF} represents the coolant feed temperature, T_C represents the coolant output temperature, C_{AF} is the concentration of the reactant A in the feed stream, C_A is the concentration of the reactant A in the outlet stream, and h is the reactor level. In this figure, some control system symbols are used. TT, FT and LT represent the temperature transmitter, the flow transmitter and the level transmitter, respectively. TC, FC and LC represent the temperature controller, the flow controller and the level controller, respectively. More detailed descriptions can be seen in the literature [47], [48].

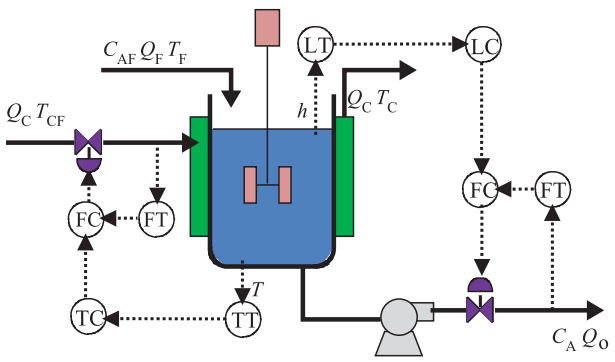


FIGURE 12. The CSTR schematic diagram.

The CSTR mechanical models involve four main differential equations, which describe the dynamic behaviors of the reactor temperature T, the reactor level h, the reactant A concentration C_A and the coolant temperature T_C. The detailed mathematical expressions can be seen in the references [48]. By simulating the CSTR mechanical models, normal and fault operation datasets are generated. The simulated six fault types can be seen in Table 3, which include the valve sticking, the sensor fault, and the process changes.

TABLE 3. The CSTR system fault patterns.

Fault	Description
F1	Coolant valve is fixed on the constant value 0.297.
F2	Feed temperature ramps up with the rate of 0.01K/min.
F3	Coolant temperature ramps down with the rate of 0.01K/min.
F4	Reactor temperature sensor has a bias of 0.15K.
F5	Catalyst activation energy ramps up with the rate of 0.5K/min.
F6	Heat-transfer coefficient ramps down with the rate of $10J/(min^2 \cdot K)$.

Each dataset includes ten variables [47] and all the measured variables are collected considering the Gaussian noises. A normal operation dataset with 300 samples is generated for statistical modeling while another 2000 normal samples are simulated to constitute the validation dataset for the confidence limit determination. Each fault dataset includes 800 samples, where the fault is injected from the 241st sample instant.

Three methods of KPCA, SLKPCA, and PA-SLKPCA are applied to detect the fault. The kernel width σ is selected as 100 and the moving window w is set to 60. For the PA-SLKPCA method, we assume that two faults F1 and F2 are prior. The monitoring results are analyzed as follows.

Firstly we investigates the fault F1 and its monitoring charts obtained by three methods are listed in Figs. 13, 14, and 15. Fault F1 involves the valve failure. The normal value of the coolant control valve is 0.3. When the fault occurs, the coolant control valve is fixed on the constant value 0.297. It is clear that the deviation is very slight. So this fault can be viewed as an incipient fault. However, this fault means the malfunction of the coolant control valve, which may lead to the serious result when some large disturbances occur because it factually affects the reactor temperature control. As the valve failure value is very close to the normal condition, this fault is difficult to detect by the traditional KPCA method. The KPCA monitoring charts are shown in Fig. 13, where the T^2 FDR is 0.030 and the Q FDR is 0.048. As SLKPCA is developed by incorporating the statistical local analysis for the incipient changes detection, it gives better performance in the Fig. 14. The SLKPCA Q statistic detects this fault with the higher FDR of 0.627 and the T^2 statistic also makes a certain improvement with the FDR of 0.245. By further considering the fault information, PA-SLKPCA gives the best performance, which has the highest FDR of 0.964 and 0.936 for the WT^2 and WQ, respectively.

Furthermore, the process monitoring charts on the fault F5 are demonstrated in Figs. 16, 17, and 18. The fault F5 involves the slow change of the catalyst activation energy parameter ER. As this fault occurs, ER ramps up with the rate of 0.5K/min. Compared with the nominal value 8750K, the fault magnitude is very small and can be thought

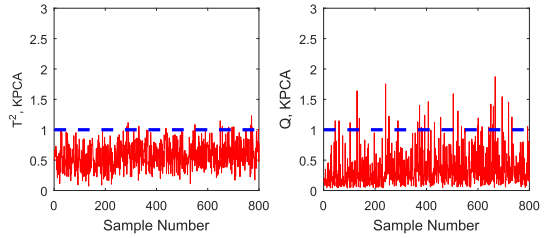


FIGURE 13. The monitoring charts of the fault F1 based on KPCA.

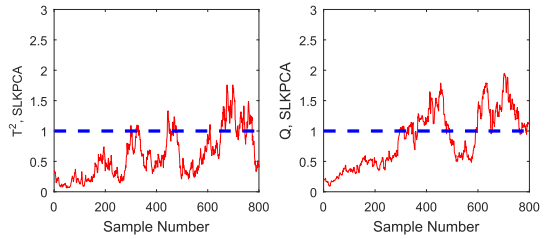


FIGURE 14. The monitoring charts of the fault F1 based on SLKPCA.

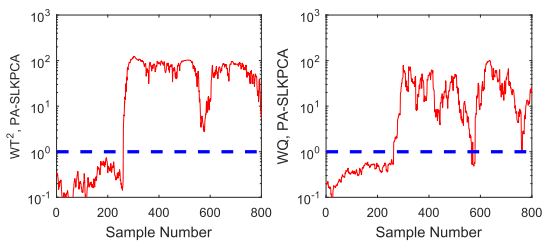


FIGURE 15. The monitoring charts of the fault F1 based on PA-SLKPCA.

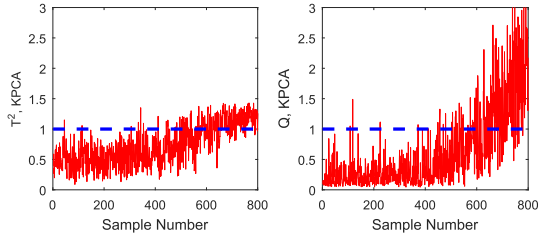


FIGURE 16. The monitoring charts of the fault F5 based on KPCA.

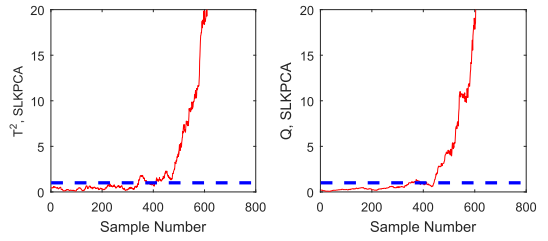


FIGURE 17. The monitoring charts of the fault F5 based on SLKPCA.

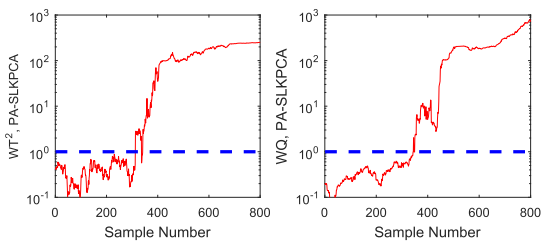


FIGURE 18. The monitoring charts of the fault F5 based on PA-SLKPCA.

TABLE 4. Fault detection rates on CSTR system faults obtained by KPCA, SLKPCA and PA-SLKPCA.

Method	KPCA		SLKPCA		PA-SLKPCA	
Statistic	T^2	Q	T^2	Q	WT^2	WQ
F1	0.030	0.048	0.245	0.627	0.964	0.936
F2	0.021	0.073	0.077	0.336	0.684	0.373
F3	0.036	0.071	0.252	0.320	0.686	0.514
F4	0.020	0.021	0.000	0.214	0.143	0.430
F5	0.361	0.313	0.763	0.714	0.863	0.813
F6	0.057	0.096	0.445	0.536	0.734	0.718
average	0.118	0.125	0.365	0.446	0.606	0.619

TABLE 5. Fault detection times on CSTR system faults obtained by KPCA, SLKPCA and PA-SLKPCA.

Method	KPCA		SLKPCA		PA-SLKPCA	
Statistic	T^2	Q	T^2	Q	WT^2	WQ
F1	-	-	660	486	371	453
F2	-	-	694	557	249	334
F3	-	-	443	296	261	263
F4	-	-	800	381	429	381
F5	584	663	343	354	313	350
F6	-	-	368	417	368	403

as one incipient fault. The increasing of ER leads to the slowing-down of the reaction, which further results in the rising of the reactant A concentration. That means the degradation of product quality. As no prior information on this fault is involved in the auxiliary model, PA-SLKPCA can not ensure the performance improvement. However, two PA-SLKPCA statistics still do better than the KPCA and SLKPCA methods. It is straightforward that KPCA can not detect this fault effectively. SLKPCA T^2 and Q detect the fault F5 at the 343rd and 354th sample, respectively. By contrast, PA-SLKPCA WT^2 detects the fault F5 at the 313rd sample, 30 samples earlier than the SLKPCA T^2 statistic, while PA-SLKPCA WQ alarms the fault at the 350th sample, also 4 samples earlier than the SLKPCA Q statistic. The FDRs of PA-SLKPCA WT^2 and WQ statistics are 0.863 and 0.813, respectively, which are higher than the corresponding SLKPCA's FDRs of 0.763 and 0.714. On the whole, PA-SLKPCA still performs satisfactory in the detection of unknown fault type.

For an overall comparison, the FDRs and FDTs of three methods are listed in the Table 4 and 5. From Table 4, the average FDRs of KPCA are 0.118 and 0.125 for T^2 and Q, respectively. SLKPCA improves the fault detection rates to 0.365 and 0.446, respectively. By incorporating the prior fault information, PA-SLKPCA further enhances the process monitoring performance with a higher average FDRs of 0.606 and 0.619 for WT^2 and WQ, respectively. The fault detection rates in Table 5 also demonstrate the superiority of the proposed PA-SLKPCA method. The average FARs (AFARs) of three methods are tabulated in the Table 6. Considering each fault testing dataset involves 240 normal samples, there are 1440 normal samples for all the six fault datasets. AFAR is computed as the ratio of the alarming samples over the total 1440 samples. By the Table 6, we observe that all the methods

TABLE 6. Average False alarming rates (FARs) on CSTR system faults obtained by KPCA, SLKPCA and PA-SLKPCA.

Method	KPCA		SLKPCA		PA-SLKPCA	
Statistic	T^2	Q	T^2	Q	WT^2	WQ
AFAR	0.009	0.010	0	0	0.008	0.007

are consistent with the 99% confidence limit, which means the AFAR is around 0.01.

In the above analysis, we only test the case with the fixed fault samples. Next we discuss the influence of the numbers of the fault samples. For the incipient faults involving the slope changes, these incipient faults may evolve into the significant faults. So if we increase the fault testing dataset number gradually, the fault detection rate will be improved obviously because the matured fault is easier to detect. However, for the incipient faults resulted from the step change of process variables, the process conditions from the incipient to mature stages are very similar so that the fault detection rates are not affected remarkably. To demonstrate this, we change the fault testing dataset number as 400, 500, 600, 700 and 800, respectively. The fault F1 and F5 are illustrated and the corresponding FDRs are plotted in Fig. 19. As to the fault F1, which belongs to the incipient fault with a small step change, it is seen that its WT^2 and WQ FDRs are not influenced seriously although they are enhanced slightly when the fault testing dataset number is enlarged. By contrast, a clear uptrend of FDR on fault F5 can be observed along with the increasing of the fault testing dataset number. That is because the fault F5 is the incipient fault involving the slope process change. Generally, the results in Fig. 19 validate our analysis.

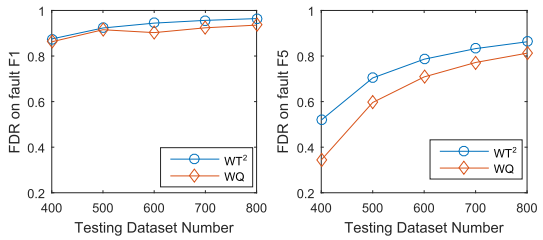


FIGURE 19. The FDRs on the fault F1 and F5 with different fault sample numbers.

VI. CONCLUSION

To make full use of all available process information is helpful to process monitoring and fault diagnosis. Based on this strategy, this paper proposes an improved SLKPCA method PA-SLKPCA for incipient fault detection. The most important contribution is the design of the primary-auxiliary monitoring framework for incipient fault detection. In addition, this paper presents a double-block modeling strategy for the development of the auxiliary SLKPCA model, and applies the Bayesian inference based weighted fusion strategy to integrate the monitoring results from the primary and auxiliary models. The applications to the numerical system

and the simulated CSTR process are performed to demonstrate the effectiveness of the proposed method. Based on the results and discussions, some conclusions can be drawn as follows. (1) If the occurred fault is same to the prior fault, the auxiliary model comes into play and PA-SLKPCA can perform better than SLKPCA significantly. (2) If the occurred fault is different to the prior faults, the auxiliary model may also help, or may not help, but it definitely does not worsen the fault detection results. (3) Generally, the PA-SLKPCA method based on the utilization of prior fault information is beneficial to the incipient fault detection. In fact, the idea of utilizing the prior process information can be extended to many other fault detection and diagnosis tasks.

Although the advantages of the proposed method have been successfully demonstrated, there are some related issues deserving further studies. One important issue is about the incipient fault identification and classification. In this paper, we only focus on detecting the incipient faults. Once one fault is detected, how to identify the sources and classify its type is the following task. Some fault identification and classification methods have been developed in the present works [49], [50], and how to extend them to the incipient fault case is one valuable research topic.

REFERENCES

[1] 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.

[2] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008–2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, 2018.

[3] S. Yin, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 657–667, Jan. 2015.

[4] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: A review," *IEEE Access*, vol. 6, pp. 66723–66741, 2018.

[5] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, Jul. 2017.

[6] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 560–572, Mar. 2018.

[7] F. Jia, E. B. Martin, and A. J. Morris, "Non-linear principal components analysis for process fault detection," *Comput. Chem. Eng.*, vol. 22, no. S1, pp. S851–S854, 1998.

[8] H. Vedam and V. Venkatasubramanian, "PCA-SDG based process monitoring and fault diagnosis," *Control Eng. Pract.*, vol. 7, no. 7, pp. 903–917, 1999.

[9] Z. Chen, C. Yang, T. Peng, H. Dan, C. Li, and W. Gui, "A cumulative canonical correlation analysis-based sensor precision degradation detection method," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6321–6330, Aug. 2019.

[10] C. Tong, T. Lan, H. Yu, and X. Peng, "Distributed partial least squares based residual generation for statistical process monitoring," *J. Process Control*, vol. 75, pp. 77–85, Mar. 2019.

[11] Q. Jiang and X. Yan, "Parallel PCA–KPCA for nonlinear process monitoring," *Control Eng. Pract.*, vol. 80, pp. 17–25, Nov. 2018.

[12] L. Zhou, J. Chen, B. Hou, and Z. Song, "Multi-grade principal component analysis for fault detection with multiple production grades," *Chemometrics Intell. Lab. Syst.*, vol. 175, pp. 20–29, Apr. 2018.

[13] Y. Dong and S. J. Qin, "A novel dynamic PCA algorithm for dynamic data modeling and process monitoring," *J. Process Control*, vol. 67, pp. 1–11, Jul. 2018.

[14] M. Alauddin, F. Khan, S. Imtiaz, and S. Ahmed, "A bibliometric review and analysis of data-driven fault detection and diagnosis methods for process systems," *Ind. Eng. Chem. Res.*, vol. 57, no. 32, pp. 10719–10735, 2018.

[15] J. Zhu, Z. Ge, Z. Song, and F. Gao, "Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data," *Annu. Rev. Control*, vol. 46, pp. 107–133, Oct. 2018.

[16] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial big data for fault diagnosis: Taxonomy, review, and applications," *IEEE Access*, vol. 5, pp. 17368–17380, 2017.

[17] H. Chen, B. Jiang, and N. Lu, "A newly robust fault detection and diagnosis method for high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2198–2208, Jun. 2019.

[18] H. Chen and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, to be published. doi: 10.1109/TITS.2019.2897583.

[19] S. Wold, "Exponentially weighted moving principal components analysis and projections to latent structures," *Chemometrics Intell. Lab. Syst.*, vol. 23, no. 1, pp. 149–161, 1994.

[20] H. Ji, X. He, J. Shang, and D. Zhou, "Exponential smoothing reconstruction approach for incipient fault isolation," *Ind. Eng. Chem. Res.*, vol. 57, no. 18, pp. 6353–6363, 2018.

[21] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part I," *Signal Process.*, vol. 94, no. 1, pp. 278–287, 2014.

[22] H. Chen, B. Jiang, and N. Lu, "An improved incipient fault detection method based on kullback-leibler divergence," *ISA Trans.*, vol. 79, pp. 127–136, Aug. 2018.

[23] H. Chen, B. Jiang, W. Chen, and H. Yi, "Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4716–4725, Jun. 2019.

[24] C. H. Zhao and F. R. Gao, "A sparse dissimilarity analysis algorithm for incipient fault isolation with no priori fault information," *Control Eng. Pract.*, vol. 65, pp. 70–82, Aug. 2017.

[25] S. Zhang, C. Zhao, and F. Gao, "Incipient fault detection for multiphase batch processes with limited batches," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 1, pp. 103–117, Jan. 2019.

[26] W. S. Ge, J. Wang, J. L. Zhou, H. Wu, and Q. B. Jin, "Incipient fault detection based on fault extraction and residual evaluation," *Ind. Eng. Chem. Res.*, vol. 54, no. 14, pp. 3664–3677, 2015.

[27] J. Shang, D. Zhou, M. Chen, H. Ji, and H. Zhang, "Incipient sensor fault diagnosis in multimode processes using conditionally independent Bayesian learning based recursive transformed component statistical analysis," *J. Process Control*, vol. 77, pp. 7–19, May 2019.

[28] M. Basseville, "On-board component fault detection and isolation using the statistical local approach," *Automatica*, vol. 34, no. 11, pp. 1391–1415, Nov. 1998.

[29] U. Kruger, S. Kumar, and T. Littler, "Improved principal component monitoring using the local approach," *Automatica*, vol. 43, no. 9, pp. 1532–1542, Sep. 2007.

[30] Z. Chen, K. Zhang, S. X. Ding, Y. A. W. Shardt, and Z. Hu, "Improved canonical correlation analysis-based fault detection methods for industrial processes," *J. Process Control*, vol. 41, pp. 26–34, May 2016.

[31] N. Li, W. Yan, and Y. Yang, "Spatial-statistical local approach for improved manifold-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 54, no. 34, pp. 8509–8519, 2015.

[32] Z. Ge, C. Yang, and Z. Song, "Improved kernel PCA-based monitoring approach for nonlinear processes," *Chem. Eng. Sci.*, vol. 64, no. 9, pp. 2245–2255, May 2009.

[33] X. Deng and J. Deng, "Incipient fault detection for chemical processes using two-dimensional weighted SLKPCA," *Ind. Eng. Chem. Res.*, vol. 58, no. 6, pp. 2280–2295, Jan. 2019.

[34] L. Wang, "Enhanced fault detection for nonlinear processes using modified kernel partial least squares and the statistical local approach," *Can. J. Chem. Eng.*, vol. 96, no. 5, pp. 1116–1126, 2018.

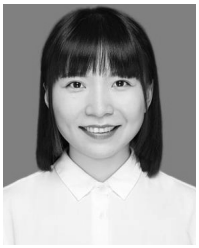
[35] J.-M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 59, no. 1, pp. 223–234, Jan. 2004.

[36] M. Navi, N. Meskin, and M. Davoodi, "Sensor fault detection and isolation of an industrial gas turbine using partial adaptive KPCA," *J. Process Control*, vol. 64, pp. 37–48, Apr. 2018.

[37] Y. Zhang, S. Li, and Y. Teng, "Dynamic processes monitoring using recursive kernel principal component analysis," *Chem. Eng. Sci.*, vol. 72, pp. 78–86, Apr. 2012.

[38] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[39] R. T. Samuel and Y. Cao, "Nonlinear process fault detection and identification using kernel PCA and kernel density estimation," *Syst. Sci. Control Eng.*, vol. 4, no. 1, pp. 165–174, Jun. 2016.
[40] J. Harmouche, C. Delpha, D. Diallo, and Y. L. Bihan, "Statistical approach for nondestructive incipient crack detection and characterization using kullback-leibler divergence," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1360–1368, Sep. 2016.
[41] N. Bouhlel and A. Dziri, "Kullback–Leibler divergence between multivariate generalized Gaussian distributions," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1021–1025, Jul. 2019.
[42] P. Wu, "Performance monitoring of MIMO control system using Kullback–Leibler divergence," *Can. J. Chem. Eng.*, vol. 96, no. 7, pp. 1559–1565, 2018.
[43] Z. Ge, M. Zhang, and Z. Song, "Nonlinear process monitoring based on linear subspace and Bayesian inference," *J. Process Control*, vol. 20, no. 5, pp. 676–688, 2010.
[44] J. Zhu, Z. Ge, Z. Song, L. Zhou, and G. Chen, "Large-scale plant-wide process modeling and hierarchical monitoring: A distributed Bayesian network approach," *J. Process Control*, vol. 65, pp. 91–106, May 2018.
[45] C. Xu, S. Zhao, and F. Liu, "Sensor fault detection and diagnosis in the presence of outliers," *Neurocomputing*, vol. 349, pp. 156–163, Jul. 2019.
[46] L. M. Elshenawy and T. A. Mahmoud, "Fault diagnosis of time-varying processes using modified reconstruction-based contributions," *J. Process Control*, vol. 70, pp. 12–23, Oct. 2018.
[47] X. Deng, N. Zhong, and L. Wang, "Nonlinear multimode industrial process fault detection using modified kernel principal component analysis," *IEEE Access*, vol. 5, pp. 23121–23132, 2017.
[48] X. Deng and X. Tian, "Sparse kernel locality preserving projection and its application in nonlinear process fault detection," *Chin. J. Chem. Eng.*, vol. 21, no. 2, pp. 163–170, 2013.
[49] R. Jia, J. Wang, and J. Zhou, "Fault diagnosis of industrial process based on the optimal parametric t-distributed stochastic neighbor embedding," *Sci. China Inf. Sci.*, to be published. [Online]. Available: http://engine.scichina.com/publisher/scp/journal/SCIS/doi/10.1007/s11432-018-9807-7?slug=abstract. doi: 10.1007/s11432-018-9807-7.
[50] X. Chen, J. Wang, and J. Zhou, "Probability density estimation and Bayesian causal analysis based fault detection and root identification," *Ind. Eng. Chem. Res.*, vol. 57, no. 43, pp. 14656–14664, 2018.

XIAOGANG DENG received the B.Eng. and Ph.D. degrees from the China University of Petroleum, Dongying, China, in 2002 and 2008, respectively. From October 2015 to October 2016, he was a Visiting Scholar with the Department of Electronics and Computer Sciences, University of Southampton, Southampton, U.K. He is currently an Associate Professor with the College of Information and Control Engineering, China University of Petroleum. His research interests include industrial process modeling and simulation, data-driven fault detection and diagnosis, and control performance monitoring.



PEIPEI CAI received the B.E. degree from Yantai University, in 2017. She is currently pursuing the master's degree with the College of Information and Control Engineering, China University of Petroleum, Qingdao. Her interests include statistical data mining and incipient fault detection.

JIAWEI DENG received the bachelor's degree in engineering from the Harbin University of Science and Technology, in 2016, and the master's degree in engineering from the China University of Petroleum, in 2019. Her research interests include machine learning theory and industrial process monitoring.

YUPING CAO received the B.Eng. degree in engineering and the Ph.D. degree in control theory and control engineering from the China University of Petroleum, in 2004 and 2010, respectively, where she is currently a Lecturer with the College of Information and Control Engineering. Her research interest includes process monitoring, fault diagnosis, and fault prediction.

ZHIHUAN SONG received the B.Eng. and M.Eng. degrees in industrial automation from the 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 a Postdoctoral Research Fellow and an Associate Professor, and currently a Professor. He has published more than 200 papers in journals and conference proceedings. His research interests include the modeling and fault diagnosis of industrial processes, embedded control systems, and advanced process control technologies.