Chemometrics and Intelligent Laboratory Systems 127 (2013) 121–131

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical process monitoring

Qingchao Jiang, Xuefeng Yan *

Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 7 March 2013
Received in revised form 19 June 2013
Accepted 19 June 2013
Available online 1 July 2013

Keywords:
Weighted kernel principal component analysis
Nonlinear process monitoring
Probability density estimation
Moving window

ABSTRACT

Kernel principal component analysis (KPCA) has been widely used in nonlinear process monitoring; however, KPCA does not always perform efficiently because useful information may be submerged under retained KPCs. To address this shortcoming, probability density estimation- and moving weighted window-based KPCA (PM-WKPCA) is proposed. PM-WKPCA is used mainly to estimate the probability and evaluate the importance of each KPC by kernel density estimation and then set different weighting values on KPCs to highlight the useful information. The status of the process is also evaluated comprehensively using weighted statistics within a moving window. The efficiency of the proposed method is demonstrated by the following: case studies on a numerical nonlinear system, the simulated continuously stirred tank reactor process, and the Tennessee Eastman process. Monitoring results indicate that the proposed method is superior to the conventional PCA, KPCA, and some typical extension methods.

1. Introduction

Process monitoring has gained increasing interest because of the rising demand in plant safety and product quality. With recent developments in data collection and computing technology, multivariate statistical process monitoring (MSPM) methods [1–14] have progressed quickly [15–17]. And among these MSPM methods, principal component analysis (PCA) usually serves as the most fundamental one and has been studied intensively and used extensively. PCA can be used effectively with high-dimensional, highly correlated data by projecting the data onto a lower dimensional subspace containing sufficient variance information of normal training data. However, PCA methods assume that the relationship between process variables is linear, which can easily change in practice [18–20].

With the nonlinear behavior of the chemical process considered, numerous nonlinear PCA (NLPCA) approaches have been developed [13,20,21], and the kernel PCA (KPCA) [22] method, which can efficiently compute PCs in a high-dimensional feature space using the kernel function, is the most widely used [22]. The key idea of KPCA is to map the data initially into a feature space by nonlinear mapping and then extract the PCs in the high-dimensional feature space. A major advantage of KPCA is that it only requires solving an eigenvalue problem and requires no nonlinear optimization [18,19]. KPCA has gained considerable attention because of its simplicity and efficiency, and it has also been extended to dynamic KPCA [23] and multiscale KPCA [24], among others, to solve various monitoring problems [10,25–27].

Despite sufficient research on KPCA and numerous successful applications, the approach does not always perform efficiently. In KPCA monitoring, the KPCA model is generated from the normal training data. The monitored variables are first nonlinearly mapped into a high-dimensional space through the kernel mapping, and then the kernel principal components (KPCs) are generated as linear combinations of the variables in the high-dimensional feature space. Generally, the first several KPCs are employed to construct the dominant subspace [18,19,28], and correspondingly, T^2 statistic is used to monitor the variation in the dominant subspace. In the T^2 statistic, the employed KPCs are scaled with similar importance degrees [15,28,29]. However, for a definite fault, the fault usually causes one or several monitored variables' variation, and all the employed KPCs are not guaranteed to have the same variation degree [30]. Namely, there are some KPCs with larger variation, and others with less or without variation. In all the employed KPCs, the KPCs with larger variation contain more fault information and are beneficial to find fault timely, and the KPCs with less variation or without variation contain less or no fault information and aren't used to find fault. Thus, for monitoring the definite fault, the importance degrees of all the employed KPCs are usually not the same (as illustrated in the Motivational example section below). Some KPCs would reflect more information of the fault while others would reflect less. When all the employed KPCs are used to monitor process with the same importance by the T^2 statistics, the useful information (contained in the KPCs with larger variation) for detecting and diagnosing the fault may be submerged by the KPCs with less variation and without variation, and

* Corresponding author at: East China University of Science and Technology, P.O. Box 293, MeiLong Road No. 130, Shanghai 200237, PR China. Tel.: +86 2164251036.
E-mail address: xfyan@ecust.edu.cn (X. Yan).

0169-7439/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chemolab.2013.06.013

poor monitoring performance happens. Therefore, the importance degree of each employed KPC should be evaluated online and the weighting strategy is needed to highlight the fault information.

Considering the issue of highlighting the useful variation information, several strategies used in MSPM from different aspects have been reported. Wold [31] proposed a timely updated model that weights recent observations more heavily than earlier ones. He et al. [32] highlighted the fault information by weighting related variables heavily in kernel Fisher discriminant analysis for fault diagnosis. Ferreira et al. [33] suggested a sample-wise weighted PCA for multicampaign process monitoring. These methods mainly focused on weighting raw measured variables without analyzing the importance of different latent variables. Rashid and Yu [10,27] used the multidimensional mutual information to evaluate the statistical dependency between the independent component subspaces of the normal benchmark and monitored data sets, and constructed a dissimilarity index which considered the online fault information. This method highlighted the fault information in the ICA dominant subspace, however, the behavior of each latent variable was not analyzed. Jiang and Yan [30] proposed weighted principal component analysis (WPCA) to highlight the useful information reflected on the dominant subspace. Online WPCA evaluates the importance of each PC by using the change rate of the T^2 statistic along each PC and setting different values on the PCs to highlight the useful information for process monitoring. This method improved both fault detection and identification performance; however, as is well known, the nonlinear relationships between variables can't be well extracted by linear PCA, and therefore the WPCA is limited to linear processes.

In this article, probability density estimation- and moving weighted window-based KPCA (PM-WKPCA) is proposed to solve the problem on submerged useful information and improve performance in nonlinear chemical process monitoring. To find and stress the informative KPCs in time, the probability density estimation (PDE) method is employed. First, the probability density function of each KPC in normal condition is estimated through kernel density estimation (KDE) with large amount of normal process data. Second, when online monitoring, the probability of each KPC score in the current sample is determined by the previously obtained distribution, and the importance of the corresponding KPC is evaluated. Then, different weighting values are adaptively and objectively set on these KPCs according to importance (P-WKPCA). Moreover, with the relationship between the previous and the current status considered, the status of the process is finally evaluated within a moving window to improve further the monitoring performance.

The rest of this article is structured as follows. First, the KPCA model used in process monitoring and KDE probability estimation are briefly reviewed, followed by a motivational example that illustrates the submerged useful information problem. Second, PM-WKPCA monitoring is proposed and some details are presented. In Section 4, the proposed monitoring method is tested in a numerical nonlinear process, the continuous stirred tank reactor (CSTR) process, and the Tennessee Eastman (TE) process. The monitoring results and several comparisons with conventional PCA, KPCA, and WPCA are presented. Finally, the conclusions in this study are given in Section 5.

2. Preliminaries

2.1. Kernel principal component analysis

In KPCA, observations are nonlinearly mapped into a high-dimensional feature space F and then linear PCA is employed to extract the nonlinear correlation between the variables [18,19,22]. Let the normalized training set be $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N \in R^m$ with N observations consisting of m measured process variables. The feature space is constructed by the nonlinear mapping: $R^m \rightarrow^{\boldsymbol{\Phi}(\bullet)} C^F$, where $\boldsymbol{\Phi}(\bullet)$ is the nonlinear mapping function [19,22]. The covariance matrix in the feature space F is calculated as

$$\boldsymbol{C}^F = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)^T \tag{1}$$

where $\sum_{j=1}^{N} \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right) = 0$ is assumed. The kernel principal component can be obtained by the eigenvalue problem [18,19,22]

$$\lambda \boldsymbol{v} = \boldsymbol{C}^F \boldsymbol{v} = \frac{1}{N}\sum_{j=1}^{N} \left\langle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)^T, \boldsymbol{v} \right\rangle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right)\lambda \boldsymbol{v} \tag{2}$$

where λ and $\mathbf{v}$ denote the eigenvalue and the eigenvector of the covariance matrix $\mathbf{C}^F$, respectively, and $\langle \boldsymbol{x},\boldsymbol{y} \rangle$ denotes the dot product between $\mathbf{x}$ and $\mathbf{y}$. This implies that all solutions $\mathbf{v}$ with $\lambda \neq 0$ must lie in the span of $\boldsymbol{\Phi}(\boldsymbol{x}_1),\ldots,\boldsymbol{\Phi}(\boldsymbol{x}_N)$. Then, $\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v}$ is equivalent to $\lambda \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{v} \rangle = \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{C}^F \boldsymbol{v} \rangle (k = 1,\ldots,N)$. There exist coefficients α_i $(i = 1,\ldots,N)$ such that $\boldsymbol{v} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\Phi}(\boldsymbol{x}_i)$. Rewriting Eq. (2), we obtain the following:

$$\lambda \sum_{i=1}^{N} \alpha_i \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{\Phi}(\boldsymbol{x}_i) \rangle = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \left\langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \sum_{j=1}^{N} \left(\boldsymbol{x}_j\right) \right\rangle \left\langle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right), \boldsymbol{\Phi}(\boldsymbol{x}_i) \right\rangle \quad (k = 1,\ldots,N). \tag{3}$$

Note that the eigenvalue problem in Eq. (3) only involves dot products of mapped shape vectors in the feature space. We define an $N \times N$ matrix $\mathbf{K}$ by $[\boldsymbol{K}]_{ij} = K_{ij} = \langle \boldsymbol{\Phi}(\boldsymbol{x}_i), \boldsymbol{\Phi}(\boldsymbol{x}_j) \rangle$. The left-hand side of Eq. (3) can then be expressed as [18,19,22]

$$\lambda \sum_{i=1}^{N} \langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \boldsymbol{\Phi}(\boldsymbol{x}_i) \rangle = \lambda \sum_{i=1}^{N} \alpha_i K_{ki} \quad (k = 1,\ldots,N) \tag{4}$$

and the right-hand side of Eq. (3) can be expressed as

$$\frac{1}{N}\sum_{i=1}^{N} \alpha_i \left\langle \boldsymbol{\Phi}(\boldsymbol{x}_k), \sum_{j=1}^{N} \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right) \right\rangle \left\langle \boldsymbol{\Phi}\left(\boldsymbol{x}_j\right), \boldsymbol{\Phi}(\boldsymbol{x}_i) \right\rangle = \frac{1}{N}\sum_{i=1}^{N} \alpha_i \sum_{j=1}^{N} K_{kj} K_{ji} \quad (k = 1,\ldots,N). \tag{5}$$

The principal component $\mathbf{t}$ of a test vector $\mathbf{x}$ is then extracted by projecting $\boldsymbol{\Phi}(\boldsymbol{x})$ onto the eigenvectors v_k in F, as follows:

$$\boldsymbol{t}_k = \langle \boldsymbol{v}_k, \boldsymbol{\Phi}(\boldsymbol{x}) \rangle = \sum_{i=1}^{N} \alpha_i^k \langle \boldsymbol{\Phi}(\boldsymbol{x}_i), \boldsymbol{\Phi}(\boldsymbol{x}) \rangle \quad (k = 1,\ldots,p) \tag{6}$$

where p is the number of the kernel principal components retained, which is empirically determined as the smallest number of the ordered eigenvalues whose cumulative sum is above 85%.

In KPCA monitoring, the T^2 and Q statistics are constructed and monitored based on the assumption that the training data have a multivariate normal distribution in the feature space [18]. The T^2 statistic is the sum of normalized squared scores, expressed as [9,19]

$$T^2 = \left[t_1,\ldots,t_p\right] \boldsymbol{\Lambda}^{-1} \left[t_1,\ldots,t_p\right]^T \tag{7}$$

where t_k is obtained from Eq. (6), and $\boldsymbol{\Lambda}^{-1}$ is the diagonal matrix of the inverse of the eigenvalues with the retained PCs.

The Q statistic is calculated by

$$Q = \left\| \boldsymbol{\Phi}(\boldsymbol{x}) - \boldsymbol{\Phi}_p(\boldsymbol{x}) \right\|^2 = \left\| \overline{\boldsymbol{\Phi}}(\boldsymbol{x}) - \overline{\boldsymbol{\Phi}_p}(\boldsymbol{x}) \right\|^2 = \sum_{j=1}^{n} t_j^2 - \sum_{j=1}^{p} t_j^2 \tag{8}$$

where n is the number of nonzero eigenvalues among the total N eigenvalues. More details on the two statistic parameters are presented in the corresponding references [8,18,19].

2.2. Probability density estimation

Kernel density estimation introduced by Rosenblatt [34] and Parzen [35] is used to compute for the probability density of KPC scores in a feature space. A univariate kernel estimator with the kernel function K is defined by [35,36]

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left\{\frac{x-x_i}{h}\right\} \tag{9}$$

where x is the data point under consideration, x_i is an observation value from the data set, h is the window width, and n is the number of observations [8,35]. The kernel function K, which determines the shape of the bumps, has a number of possible forms [36]. However, the form of the kernel function is not very important in practice, and the Gaussian kernel is the most widely used function [36]. In this study, the Gaussian kernel is employed and the kernel estimator becomes [36]

$$f(x) = \frac{1}{n}\frac{1}{h\sqrt{2\pi}}\sum_{i=1}^{n} \exp\left(-\frac{(x-x_i)^2}{2h^2}\right). \tag{10}$$

Notably, the window width h usually has a crucial influence on the performance of KDE. Sufficient studies on the choice of h have been reported. The optimal choice of h depends on several factors such as the number of data points, data distribution, and choice of kernel function. Determining an appropriate empirical value for each specific case is also suggested [36–38]. With sufficient information on the training data, KDE is usually guaranteed to perform satisfactorily because the training data on normal status can be easily obtained.

2.3. Motivational example

In KPCA monitoring, the diagonal matrix $\boldsymbol{\Lambda}^{-1}$ in the T^2 statistic, which contains the multiplicative inverse of the estimated variances of the KPCs, scales the KPCs to the similar importance. However, for a definite fault, the importance of the KPCs is usually not the same. Some KPCs would reflect more information of the fault while others would reflect less. If the fault is a local fault and only influences very few of the KPCs or the fault information is relatively limited, the behavior of the fault may not be well reflected by using all the KPCs. The useful information for detecting and diagnosing the fault may be submerged, leading to poor monitoring performance.

To analyze the monitoring performance of KPCA and illustrate the situation in which useful information is submerged, the numerical nonlinear process involving three variables but with only one factor is considered, as originally suggested by Dong and McAvoy [21].

$$x_1 = r + e1 \tag{11}$$

$$x_2 = r^2 - 3r + e2 \tag{12}$$

$$x_3 = -r^3 + 3r^2 + e3 \tag{13}$$

where $e1$, $e2$, and $e3$ are independent noise variables $N(0,0.01)$ and $r \in [0.01,2]$. Normal data comprising 100 samples are generated according to these equations. Two sets of test data comprising 300 samples each are also generated. The following two faults are applied separately during the generation of test data sets [21].

Fault 1 A step change in x_2 by -0.3 is introduced from sample 101.
Fault 2 x_1 is linearly increased from sample 101 by adding $0.01 \times (s - 100)$ to the x_1 value of each sample in this range, where s is the sample number.

In building a conventional KPCA model to monitor the numerical process, the first three kernel principal components occupying more than 85% of the cumulative variance contribution are employed. The monitoring results of the two faults are plotted in Fig. 1, which shows that Fault 1 can be detected by the T^2 statistic in KPCA monitoring; however, the value of T^2 does not stay above the threshold, leading to a high missed detection rate and poor monitoring performance. To investigate the cause of the failure, the T^2 statistic along each kernel principal component is examined, which is calculated as

$$T_k^2 = t_k \Lambda_k^{-1} t_k^T \quad (k = 1, \ldots, p) \tag{14}$$

where Λ_k^{-1} is the (k,k)-th element in the diagonal matrix $\boldsymbol{\Lambda}^{-1}$. The monitoring results along each component are presented in Fig. 2. It can be seen from Fig. 2 that the monitoring results along different components are not the same. The T^2 statistic along the third principal component has shown larger changes when the fault occurs, indicating that most fault information is reflected on the third component. However, when the statistics are aggregated into a single T^2, the fault information in the third component would be submerged under all the components. The submerged useful information would lead to poor monitoring performance. It can be seen from Figs. 1 and 2 that directly monitoring the T_3^2 can give better monitoring results than monitoring the aggregated T^2. Therefore, the useful information should be highlighted. Considering that the fault information has no definite mapping on a certain KPC, the weighting values should be determined based on an adaptive strategy.

3. Probability density estimation and moving window-based WKPCA

PM-WKPCA for nonlinear chemical process monitoring is presented in detail.

In KPCA monitoring, the T^2 statistic is constructed to monitor the variations of the first several kernel principal components. The variations along the KPCs have been similarly scaled. However, fault information shows no definite mapping to a certain KPC, and the KPCs are not of equal importance. If the information is relatively limited, the useful KPCs could be suppressed by the useless ones, thereby submerging the useful information. Therefore, highlighting the useful KPCs is necessary for the weighting method to be applicable.

Given the weighting matrix $\mathbf{W}$, the weighted principal component $\boldsymbol{t}_W$ becomes

$$\boldsymbol{t}_W = \boldsymbol{W}\boldsymbol{t} \tag{15}$$

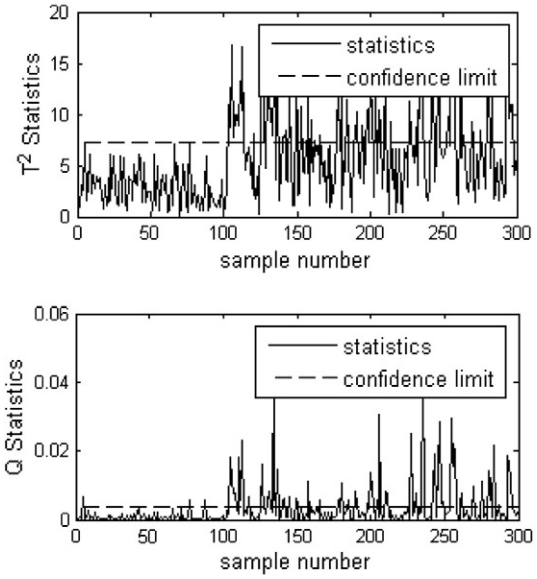

Fig. 1. Monitoring performance for Fault 1 using KPCA.

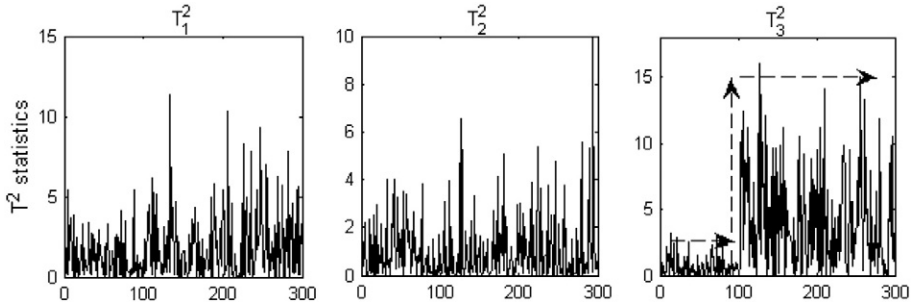

Fig. 2. Monitoring performance for Fault 1 along each kernel component.

where the weighting matrix $\boldsymbol{W} = diag(w_1,w_2,...,w_p)$ is a diagonal matrix.

The weighted T^2 statistic T_W^2 then becomes

$$T_W^2 = \left[t_1,...,t_p\right]\boldsymbol{W}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{W}\left[t_1,...,t_p\right]^T. \tag{16}$$

Given that both the weighting matrix and the $\boldsymbol{\Lambda}$ matrix are diagonal matrices, we obtain

$$T_W^2 = \left[t_1,...,t_p\right]\boldsymbol{W}^2\boldsymbol{\Lambda}^{-1}\left[t_1,...,t_p\right]^T \tag{17}$$

where $\boldsymbol{W}^2 = diag(w_1^2,w_2^2,...,w_p^2)$ remains a diagonal matrix.

The following task is to determine the weighting values in the weighting matrix **W**. Important kernel principal components should be heavily weighted, and large corresponding weighting values should be assigned. The probability density of each kernel principal component score of the training data set can be estimated based on a normal benchmark data by using KDE. When utilized for online monitoring, the current sampled point is first projected onto the kernel feature space, and the probability of the corresponding KPCs can be evaluated subsequently based on the obtained probability density function of each KPC. Considering that different KPCs provide different reflections of deviation from the reference kernel feature space, a real-time and dynamic weighting matrix $\mathbf{W}^K$ can be defined for each kernel principal component score of the currently monitored status. The purpose is to emphasize the important KPCs that represent the main deviation between the normal status and the current status. In this study, the probability function of each KPC is determined according to KDE, and only two values are considered: a larger α and a smaller β

$$w_i^K(k) = \begin{cases} 1/\alpha & \text{if } p(\overline{t}_i(k-1)) > \beta \\ 1/\beta & \text{if } p(\overline{t}_i(k-1)) \le \beta \end{cases} \tag{18}$$

where $\overline{t}_i(k-1)$ is the mean value of the previous several scores of the i-th KPC, defined as $\overline{t}_i(k-1) = \frac{1}{d}\sum_{j=k-d}^{k-1} t_i(j)$ and d is the number of samples

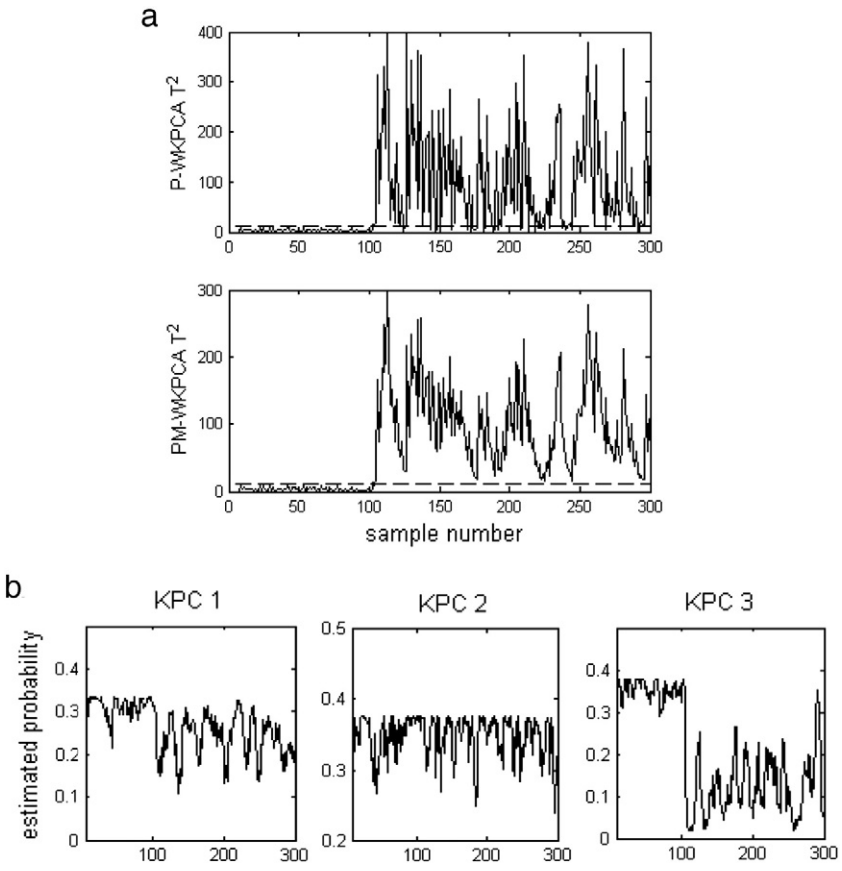

Fig. 3. Monitoring performance for Fault 1 using P-WKPCA, and PM-WKPCA; (a) P-WKPCA, and PM-WKPCA monitoring; (b) estimated probability along each KPC.

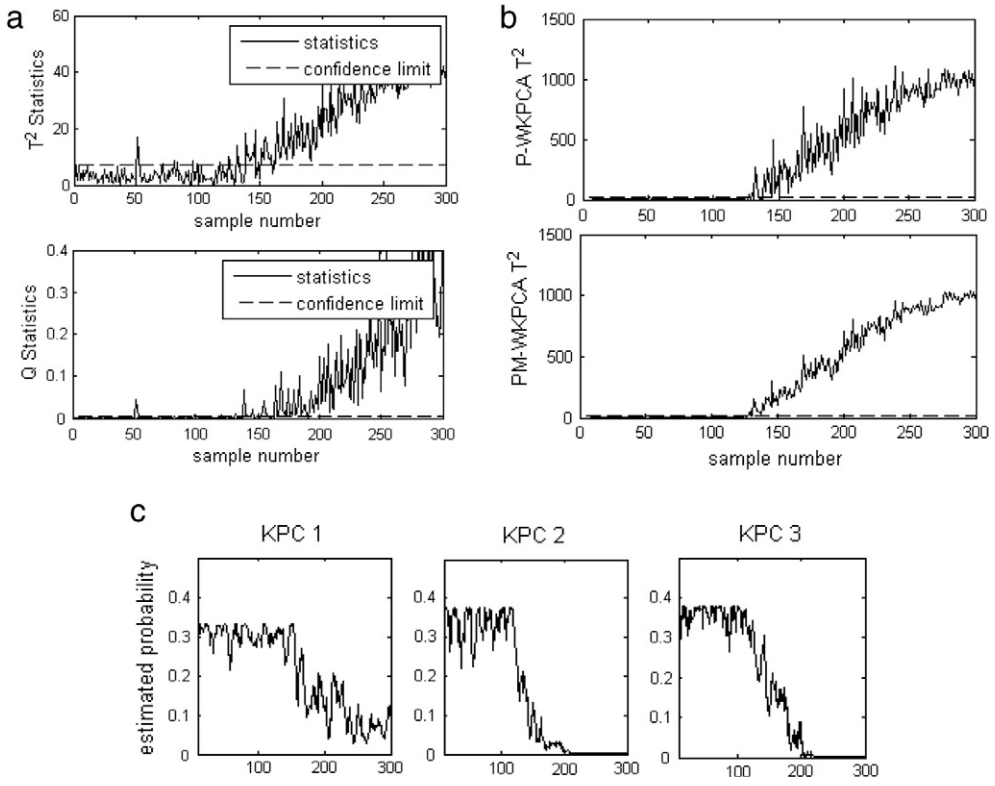

Fig. 4. Monitoring performance for Fault 2 using KPCA, P-WKPCA, and PM-WKPCA; (a) KPCA monitoring; (b) P-WKPCA, and PM-WKPCA monitoring; (c) estimated probability along each KPC.

considered; $p(\overline{t_i}(k-1))$ is the probability estimated online by KDE. This is done by using the information of the previous several points to evaluate the importance and determine the weighting values of each KPC. α and β are probability thresholds determined by an empirical study. In this study, the suggested values for α and β are 0.8–1 and 0.01–0.2, respectively.

The useful information is highlighted after probability density weighting. A moving window strategy is proposed to take the time series information into account and overcome drastic fluctuations in the monitoring result, which can further improve the performance of the T^2 statistic. Within a moving window, the T^2 statistic is defined as a comprehensive evaluation of a series of data described as

$$T^2(k) = \begin{cases} T^2(k) & \text{if } T^2(k-1) \le CL \\ \left(T^2(k), T^2(k-1), \ldots, T^2(k-n+2), T^2(k-n+1)\right)\boldsymbol{W}^S & \text{if } T^2(k-1) > CL \end{cases} \quad (19)$$

where $\boldsymbol{W}^S = [w_1^S, w_2^S, \ldots, w_n^S]^T$ is the weighting matrix functioning on the time series statistics, n is the number of points in the moving window, and CL is the confidence limit through KDE, determined using the normal process data. The elements w_i^S in $\boldsymbol{W}^S$ are defined as

$$w_i^S = \begin{cases} 1/2^i & (i = 1, \ldots, n-2) \\ 1/2^{i-1} & (i = n-1) \end{cases} \quad (20)$$

where n is the window width under consideration; therefore, $\sum_{i=1}^{n} w_i^S = 1$. The weighting values evidently decrease as time progresses, and recent samples are weighted heavily. This method addresses the current process information and properly considers the recent process history information.

The confidence limit of the weighted T^2 statistic is determined by kernel density estimation. The univariate kernel density estimator is

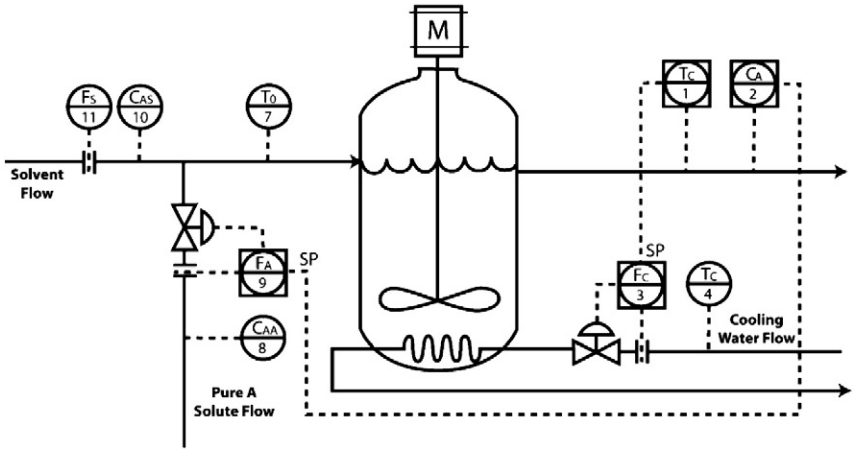

Fig. 5. Diagram of the CSTR process.

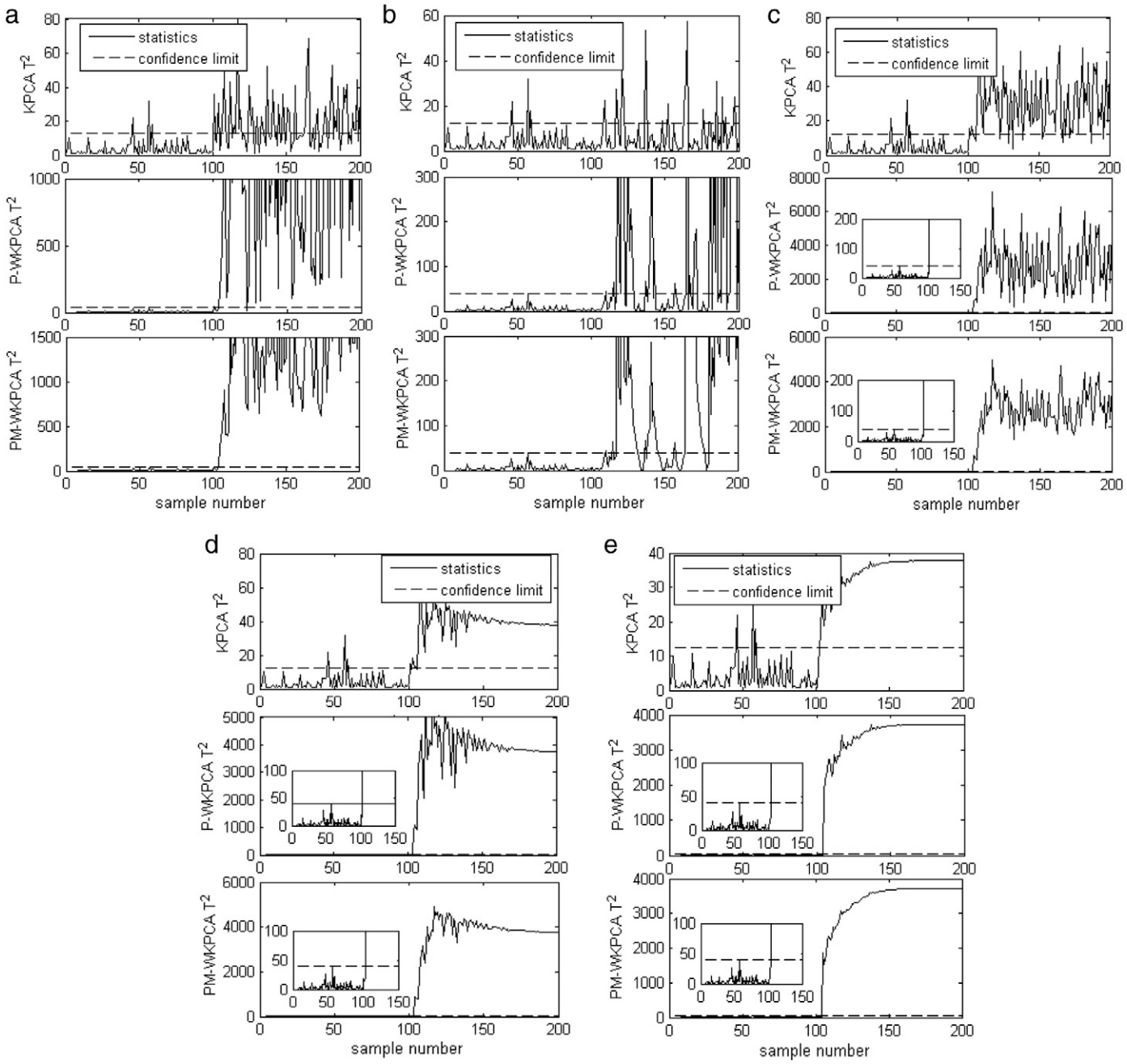

Fig. 6. Monitoring performance for faults in CSTR using KPCA, P-WKPCA, and PM-WKPCA; (a) Fault 1; (b) Fault 2; (c) Fault 3; (d) Fault 4; and (e) Fault 5.

used to estimate the density function of the normal T^2 values by using normal data. The value occupying 99% of the area of the density function can be obtained and becomes the control limit of normal operating data. A more detailed description of KDE is discussed in the studies by Parzen [35] and Lee et al. [8].

In PM-WKPCA, the first weighting strategy is to emphasize some special KPCs that capture dominant variations of current samples relative to the normal status. The second moving windows measurement is for the smooth adjustment of the weighted T^2 statistic to further improve performance monitoring. Complete monitoring procedures consist of offline and online monitoring. The detailed steps are listed below.

(1) Offline modeling

Step 1 Two normalized data sets $\mathbf{X}_A$ and $\mathbf{X}_B$ are collected under the normal status from the industrial process;

Step 2 A KPCA model is constructed using $\mathbf{X}_A$ and then $\mathbf{X}_B$ is projected onto the kernel subspace;

Step 3 KDE is used to estimate the probability density function of each KPC based on the KPC scores of $\mathbf{X}_A$;

Step 4 The probability of the KPC scores of each sample is estimated using the obtained probability density functions in Step 3, and the weighted T^2 statistics of each sample in $\mathbf{X}_B$ are calculated;

Step 5 KDE is used to determine the 99% or 95% confidence limit based on the weighted T^2 statistics of $\mathbf{X}_B$.

(2) Online monitoring

Step 1 At the current time k, the current measurement data $\mathbf{x}(k)$ is collected from the industrial process. The currently sampled data $\mathbf{x}(k)$ are normalized similar to those in the modeling steps;

Step 2 The KPCs of $\mathbf{x}(k)$ are calculated;

Step 3 The probability of each KPC is estimated, and the real-time and dynamic weighting matrix $\mathbf{W}^K(k)$ is determined;

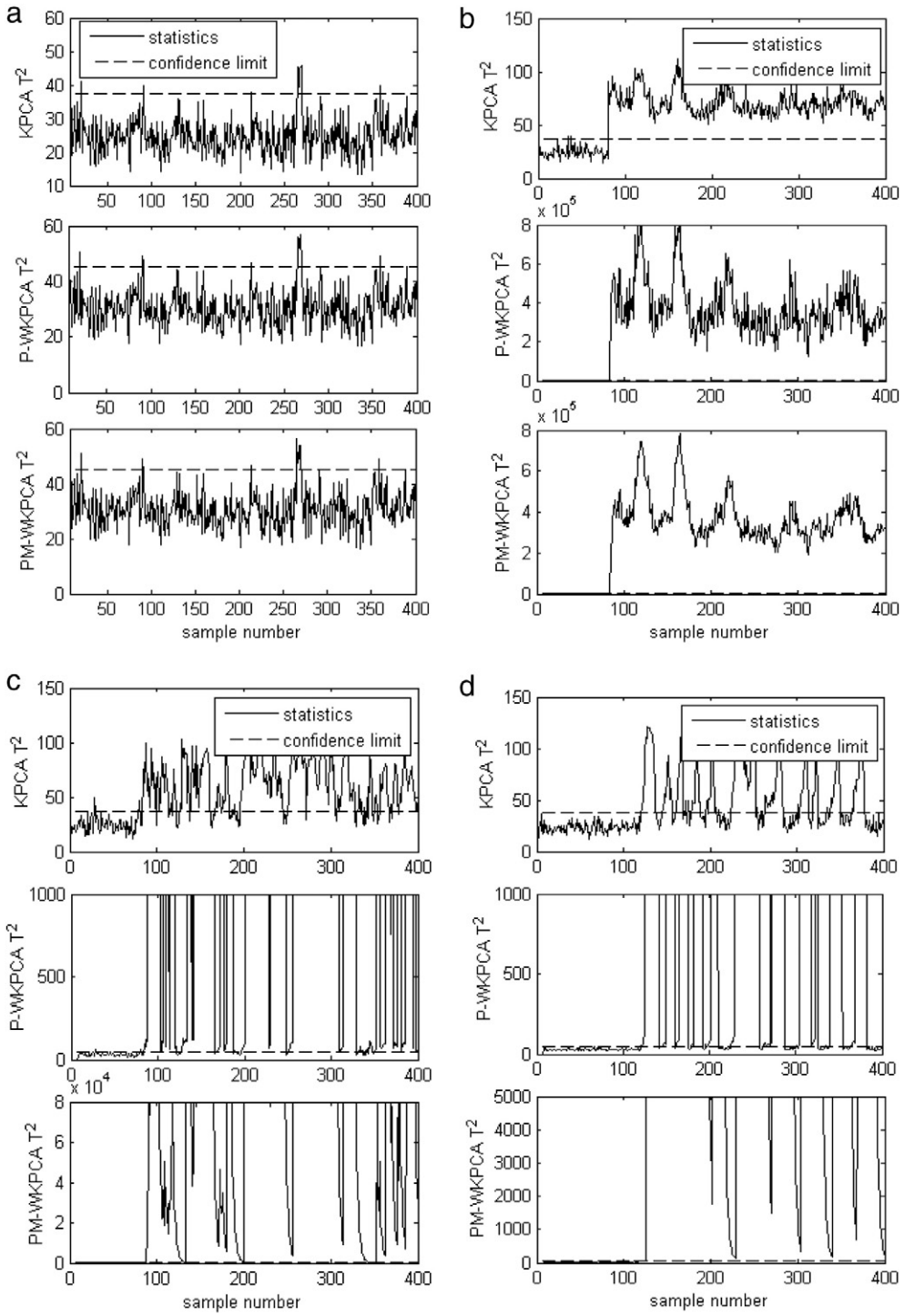

Fig. 7. Monitoring performance for faults in the TE process by KPCA, P-WKPCA, and PM-WKPCA; (a) Fault 0; (b) Fault 4; (c) Fault 11; and (d) Fault 20.

Step 4 The $T^2(k)$ value for the current sample data is calculated, and the weighted statistic $T^2(k)$ in the moving window is then calculated;

Step 5 If the $T^2(k)$ statistic exceeds the confidence limit, detection of certain fault types is indicated; fault diagnosis tools are then used to analyze the root cause. Otherwise, Step 1 is repeated and monitoring continues.

4. Case studies

In this section, the proposed PM-WKPCA monitoring method is applied in the numerical nonlinear, CSTR, and TE benchmark processes. The missed detection (Type I error) indicates the condition that the point is fault points but the statistics fail to detect it, while the false alarm (Type II error) indicates the condition that there is no fault, but the statistics exceed the limits.

4.1. Case study of the numerical nonlinear process

First, the numerical nonlinear process discussed in the motivational example is considered. In building the weighted KPCA model, the probability threshold β is set as 0.2, and the moving window width n is an empirical value 5. The monitoring results of the two faults which have been analyzed in the Motivational

example section are plotted in Fig. 1. The monitoring results using P-WKPCA and PM-WKPCA are presented in Fig. 3(a) which indicate the number of missed detection points being significantly reduced. PM-WKPCA is shown to perform most efficiently among the three methods. The estimated probabilities of the KPCs are presented in Fig. 3(b) for a visualized analysis of the necessity for the weighting strategy. In Fig. 3(b) the probability of the third KPC varies markedly when the fault occurs, indicating that most deviation in information is captured by the third KPC and that the KPC should be weighted with a higher value. The monitoring results of KPCA, P-WKPCA, and PM-WKPCA for Fault 2 are plotted in Fig. 4. Fig. 4(a) indicates the monitoring performance using KPCA. The T^2 monitoring results of Fault 2 are also unsatisfactory at the beginning of the fault. The monitoring results have been improved by introducing the weighting strategies, as shown in Fig. 4(b). Fig. 4(c) shows the estimated probabilities of the KPCs when monitoring Fault 2. The three KPCs are important, and the information reflected on the KPCs should be highlighted.

4.2. Case study of the CSTR process

The simulation model parameters and the simulation conditions used in the study by Yoon and MacGregor [39] are adopted in this case study. Fig. 5 presents the diagram of the process. This reactor is based on three assumptions: perfect mixing, constant physical properties, and negligible shaft work. The simulation is performed according to the following model:

$$\frac{dC_A}{dt} = \frac{F}{V}C_{A0} - \frac{F}{V}C_A - k_0 e^{-E/(RT)}C_A \tag{21}$$

$$V\rho c_p \frac{dT}{dt} = \rho c_p F(T_0 - T) - \frac{aF_C^{b+1}}{F_C + aF_C^b/\left(2\rho_C c_{pc}\right)}\left(T - T_{C,in}\right) + (-\Delta H_{rxn})Vk_0 e^{-E/(RT)}C_A. \tag{22}$$

More details on the description of the process and the parameters could be seen in Reference [39]. The process is monitored by measuring the cooling water temperature T_C, the inlet temperature T_0, the inlet concentrations C_{AA} and C_{AS}, the solvent flow F_S, the cooling water flow F_C, the outlet concentration C_A, the temperature T, and the reactant flow F_A. These nine variables form the measurement vector [40,41]

$$\boldsymbol{x} = [T_C \;\; T_0 \;\; C_{AA} \;\; C_{AS} \;\; F_S \;\; F_C \;\; C_A \;\; T \;\; F_A]^T. \tag{23}$$

The variables are sampled every minute, and 500 samples taken under normal conditions are used as the training set. In building a conventional KPCA model, the first four kernel principal components occupying more than 85% of the cumulative variance contribution are employed. Five different faults are introduced into the system and for each of the five scenarios, 200 observations are simulated. The fault is introduced at the 101th measurement. The first simulated fault, Fault 1, is a bias in the sensor of the output temperature T; the bias magnitude is 0.2 (K) (the bias magnitude is 1 (K) in the reference, and the magnitude has been reduced by 80% for acid testing). With T as a controlled variable, the effect of the fault is removed by the proportion integration controller, propagating to other variables. This fault is considered a complex fault because it affects several variables [40,41]. The monitoring results of the fault using KPCA, P-WKPCA, and PM-WKPCA are presented in Fig. 6(a), which shows that the number of missed detection points using KPCA is large. Both P-WKPCA and PM-WKPCA significantly improved the monitoring performance, with the latter performing more efficiently than the former.

The second and third faults, Faults 2 and 3, are biases in the sensors of the inlet temperature T_0 and inlet reactant concentration C_{AA}, respectively. The bias magnitude for T_0 is 0.3 (K) (reduced by 80% compared with the reference) and that for C_{AA} is 1.0 (kmol/m^3). These faults are considered simple faults because they only affect one variable. The fourth fault, Fault 4, is a drift in the sensor of C_{AA}, and its magnitude is $dC_{AA}/dt = 0.04$ (kmol/(m^3 min)) (reduced by 80% compared with the reference), which is also a simple fault. The last fault, Fault 5, is a slow drift in the reaction kinetics. The fault has the form of an exponential degradation of the reaction rate caused by catalyst poisoning. In this case, the reaction rate coefficient changes with time as $k_0(t+1) = 0.996 \times k_0(t)$. This process fault is a complex fault affecting several variables such as the output temperature T, concentration C_A, and the cooling water F_C [39,40]. The monitoring results of the faults are presented in Fig. 6(b) to (e). The figures indicate that P-WKPCA reduced the missed detection rates and improved the monitoring performance. PM-WKPCA performs most efficiently among the three monitoring methods.

4.3. Case study of the TE benchmark process

The TE process is a benchmark case in process engineering, which was developed by Downs and Vogel [42]. This case consists of five major unit operations: reactor, product condenser, vapor–liquid separator, recycle compressor, and product stripper. Two products are produced by two simultaneous gas–liquid exothermic reactions, and a byproduct is generated by two additional exothermic reactions. The process includes 12 manipulated variables, 22 continuous process measurements, and 19 compositions, as listed in Appendix Tables 1, 2, and 3. The fault simulator can generate 21 different types of faults, as shown in Table 4 of the Appendix A. All process measurements are contaminated by Gaussian noise. Once a fault enters the process, it affects almost all state variables in the process. The base control scheme for the TE process is shown in Appendix Fig. 1, and the simulation code for the open loop can be downloaded from http://brahms.scs.uiuc.edu [15,43]. The second plant-wide control structure described in the study by Lyman and Georgakis [44] is implemented to simulate the realistic conditions (closed-loop). A normal process data set (500 samples) has been collected under the base operation to develop the monitoring models. A set of 21 programmed faults (default values) are simulated, and the corresponding process data are collected for testing.

The following simulations are run in Matlab® (R2011b). A total of 33 variables (22 continuous measurements and 11 manipulated variables

Table 1
Missed detection rates of PCA, WPCA, KPCA, KICA, P-WKPCA, and PM-WKPCA.

Fault no.	PCA T^2	WPCA T^2	KPCA T^2	KICA T^2	P-WKPCA T^2	PM-WKPCA T^2
1	0.008	0.006	0	0.005	0	0
2	0.020	0.016	0.009	0.014	0.009	0.009
3	0.998	0.968	0.913	0.964	0.881	0.769
4	0.456	0.084	0	0.186	0	0
5	0.775	0.719	0.628	0.724	0.597	0.481
6	0.011	0.008	0	0.004	0	0
7	0.085	0	0	0	0	0
8	0.034	0.030	0.013	0.024	0.013	0.009
9	0.994	0.945	0.963	0.989	0.947	0.847
10	0.666	0.490	0.494	0.185	0.391	0.231
11	0.794	0.349	0.178	0.328	0.091	0.009
12	0.029	0.013	0.003	0.006	0.003	0.003
13	0.060	0.045	0.059	0.053	0.059	0.059
14	0.158	0	0	0.001	0	0
15	0.988	0.873	0.928	0.981	0.919	0.872
16	0.834	0.615	0.763	0.136	0.688	0.503
17	0.259	0.155	0.069	0.065	0.041	0.038
18	0.113	0.101	0.122	0.109	0.122	0.122
19	0.996	0.859	0.847	0.229	0.825	0.831
20	0.701	0.478	0.491	0.459	0.366	0.122
21	0.736	0.624	0.616	0.584	0.556	0.469

without the agitation speed because they were not manipulated) of the TE process are monitored in this case study. In building a conventional KPCA model, the first 23 kernel principal components occupying more than 85% of the cumulative variance contribution are employed. All faults are introduced into the process on the 81st time point. Four typical testing sets, Faults 0, 4, 11, and 20 are employed, and the monitoring results are presented. Fault 0 represents the normal operating condition in the TE process, which is usually used for testing the false alarm rates of monitoring schemes [15]. The monitoring performances of conventional KPCA, P-WKPCA, and PM-WKPCA for Fault 0 are presented in Fig. 7(a). In the figures, the monitoring performance of the normal process has not been degraded by introducing the KDE weighting and moving window strategy. The false alarm rates of KPCA, P-WKPCA, and PM-WKPCA are 0.0275, 0.0300, and 0.0275, respectively, which can be negligible in engineering practice.

Fault 4 involves a sudden temperature increase in the reactor, and the change is compensated by the control loops [15,43]. The monitoring results of Fault 4 using KPCA, P-WKPCA, and PM-WKPCA are presented in Fig. 7(b), which show prompt fault detection using the three methods. However, compared with KPCA, P-WKPCA and PM-WKPCA provided more significant discrimination between normal and abnormal observations.

When Fault 11 occurs, a random variation is introduced in the reactor cooling water inlet temperature, and large oscillations are involved in the reactor cooling water flow rate. As a result, a fluctuation exists in the reactor temperature [15,43]. The monitoring results of Fault 11 using KPCA, P-WKPCA, and PM-WKPCA are presented in Fig. 7(c). As indicated in the figure, some points under the confidence limit lead to a high missed detection rate when using KPCA. The number of missed detection points is significantly reduced when using P-WKPCA. However, the missed detection rates are further reduced, and the monitoring performance is improved when using PM-WKPCA.

Fault 20 is an unknown fault in the TE process, and the monitoring results are presented in Fig. 7(d). In the figure, the results are similar to those in monitoring Fault 11. P-WKPCA performs more efficiently than the conventional KPCA, and PM-WKPCA is the most efficient method among the three.

The missed detection rates for each fault in the TE process using PCA [15], WPCA [30], KPCA [22], KICA [45], P-WKPCA, and PM-WKPCA are summarized in Table 1. Faults 3, 9, and 15 are not used for comparison because no observable change in the mean or the variance can be detected by visually comparing the plots of each observation variable associated with Faults 3, 9, and 15 with plots associated with the normal condition. In Table 1, the use of P-WKPCA can efficiently reduce missed rates when monitoring the faults. PM-WKPCA was determined to perform most efficiently based on the comparison of the listed monitoring results.

5. Conclusions

A PM-WKPCA method has been proposed in this paper to improve the nonlinear chemical process monitoring performance. KDE is first used to calculate the probability of KPCs of online sampled data, and the varying importance of each KPC is then evaluated. Second, different weighting values are given to corresponding KPCs to highlight the KPCs that represent the main deviation information between the normal status and the current status. The process status is then evaluated comprehensively within a moving window. Compared with conventional monitoring methods, PM-WKPCA highlights the KPCs representing the more useful information for online monitoring. The technique also provides a comprehensive evaluation of the process status. As demonstrated in the case studies, the process monitoring performance is improved significantly.

Acknowledgments

The authors gratefully acknowledge the support from the following foundations: 973 project of China (2013CB733600), National Natural Science Foundation of China (21176073), Program for New Century Excel-

Appendix A

Table 1
The control variable of the TE process [15,43].

Variable no.	State	Variable no.	State
3M (1)	D feed flow (stream 2)	ME (7)	Separator pot liquid flow (stream 10)
3M (2)	E feed flow (stream 3)	XMV (8)	Stripper liquid product flow (stream 11)
XMV (3)	A feed flow (stream 1)	XMV (9)	Stripper steam valve
XMV (4)	A and C feed flow (stream 4)	XMV (10)	Reactor cooling water valve
XMV (5)	Compressor recycle value	XMV (11)	Condenser cooling water flow
XMV (6)	Purge valve (stream 9)	XMV (12)	Stirring rate

Table 2
The process variable of the TE process [15,43].

Variable no.	Process measurements	Unit
XMEAS (1)	A feed (stream 1)	km^3/h
XMEAS (2)	D feed (stream 2)	kg/h
XMEAS (3)	E feed (stream 3)	kg/h
XMEAS (4)	A and C feed (stream 4)	km^3/h
XMEAS (5)	Recycle flow (stream 8)	km^3/h
XMEAS (6)	Reactor feed rate (stream 6)	km^3/h
XMEAS (7)	Reactor pressure	kPa
XMEAS (8)	Reactor level	%
XMEAS (9)	Reactor temperature	°C
XMEAS (10)	Purge rate (stream 9)	km^3/h
XMEAS (11)	Product separator temperature	°C
XMEAS (12)	Product separator level	%
XMEAS (13)	Product separator pressure	kPa
XMEAS (14)	Product separator underflow	m^3/h
XMEAS (15)	Stripper level	%
XMEAS (16)	Stripper pressure	kPa
XMEAS (17)	Stripper underflow (stream 11)	m^3/h
XMEAS (18)	Stripper temperature	°C
XMEAS (19)	Stripper steam flow	kg/h
XMEAS (20)	Compress work	kW
XMEAS (21)	Reactor cooling water outlet temp	°C
XMEAS (22)	Separator cooling water outlet temp	°C

Table 3
The process variable of TE process-2 [15,43].

Variable no.	State	Stream no.	Sample time/min
XMEAS (23)	Composition A	6	6
XMEAS (24)	Composition B	6	6
XMEAS (25)	Composition C	6	6
XMEAS (26)	Composition D	6	6
XMEAS (27)	Composition E	6	6
XMEAS (28)	Composition F	6	6
XMEAS (29)	Composition A	9	6
XMEAS (30)	Composition B	9	6
XMEAS (31)	Composition C	9	6
XMEAS (32)	Composition D	9	6
XMEAS (33)	Composition E	9	6
XMEAS (34)	Composition F	9	6
XMEAS (35)	Composition G	9	6
XMEAS (36)	Composition H	9	6
XMEAS (37)	Composition D	11	15
XMEAS (38)	Composition E	11	15
XMEAS (39)	Composition F	11	15
XMEAS (40)	Composition G	11	15
XMEAS (41)	Composition H	11	15

Table 4
Process faults for the Tennessee Eastman process [15,43].

Fault no.	Disturbance state	Type
IDV (1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV (2)	B composition, A/C ratio constant (stream 4)	Step
IDV (3)	D feed temperature (stream 2)	Step
IDV (4)	Reactor cooling water inlet temperature	Step
IDV (5)	Condenser cooling water inlet temperature	Step
IDV (6)	A feed loss (stream 1)	Step
IDV (7)	C header pressure loss-reduced availability (stream 4)	Step
IDV (8)	A, B, C feed composition (stream 4)	Random variation
IDV (9)	D feed temperature (stream 2)	Random variation
IDV (10)	C feed temperature (stream 4)	Random variation
IDV (11)	Reactor cooling water inlet temperature	Random variation
IDV (12)	Condenser cooling water inlet temperature	Random variation
IDV (13)	Reaction kinetics	Slow drift
IDV (14)	Reactor cooling water valve	Sticking
IDV (15)	Condenser cooling water valve	Sticking
IDV (16)	Unknown	Unknown
IDV (17)	Unknown	Unknown
IDV (18)	Unknown	Unknown
IDV (19)	Unknown	Unknown
IDV (20)	Unknown	Unknown
IDV (21)	The valve for stream 4 was fixed at the steady state position	Step constant position

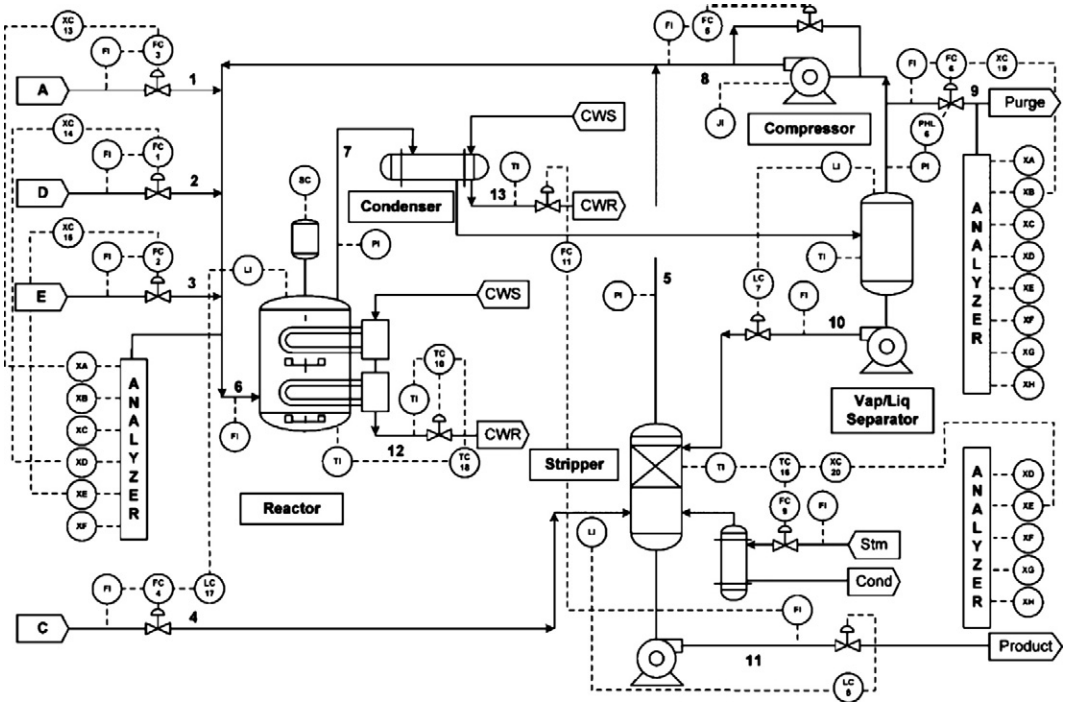

Fig. 1. Control scheme for the Tennessee Eastman process [15,43].

lent Talents in University (NCET-09-0346) and the Fundamental Research Funds for the Central Universities.

References

[1] J.E. Jackson, Quality control methods for several related variables, Technometrics 1 (1959) 359–377.
[2] T. Kourti, J.F. MacGregor, Multivariate SPC methods for process and product monitoring, J. Qual. Technol. 28 (1996) 409–428.
[3] J.V. Kresta, J.F. Macgregor, T.E. Marlin, Multivariate statistical monitoring of process operating performance, Can. J. Chem. Eng. 69 (2009) 35–47.
[4] W. Ku, R.H. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, Chemom. Intell. Lab. Syst. 30 (1995) 179–196.
[5] N. Lu, F. Wang, F. Gao, Combination method of principal component and wavelet analysis for multivariate process monitoring and fault diagnosis, Ind. Eng. Chem. Res. 42 (2003) 4198–4207.
[6] P. Nomikos, J.F. MacGregor, Monitoring batch processes using multiway principal component analysis, AICHE J. 40 (1994) 1361–1375.
[7] L.M. Elshenawy, S. Yin, A.S. Naik, S.X. Ding, Efficient recursive principal component analysis algorithms for process monitoring, Ind. Eng. Chem. Res. 49 (2009) 252–259.
[8] J.-M. Lee, C. Yoo, I.-B. Lee, Statistical process monitoring with independent component analysis, J. Process Control 14 (2004) 467–485.
[9] J.-M. Lee, C. Yoo, S.W. Choi, P.A. Vanrolleghem, I.-B. Lee, Nonlinear process monitoring using kernel principal component analysis, Chem. Eng. Sci. 59 (2004) 223–234.
[10] M.M. Rashid, J. Yu, A new dissimilarity method integrating multidimensional mutual information and independent component analysis for non-Gaussian dynamic process monitoring, Chemom. Intell. Lab. Syst. 115 (2012) 44–58.
[11] C.K. Yoo, J.-M. Lee, P.A. Vanrolleghem, I.-B. Lee, On-line monitoring of batch processes using multiway independent component analysis, Chemom. Intell. Lab. Syst. 71 (2004) 151–163.
[12] S. Joe Qin, Recursive PLS algorithms for adaptive data modeling, Comput. Chem. Eng. 22 (1998) 503–514.
[13] U. Kruger, G. Dimitriadis, Diagnosis of process faults in chemical systems using a local partial least squares approach, AICHE J. 54 (2008) 2581–2596.

[14] R. Muradore, P. Fiorini, A PLS-based statistical approach for fault detection and isolation of robotic manipulators, IEEE Trans. Ind. Electron. 59 (2012) 3167–3175.
[15] L.H. Chiang, E. Russell, R.D. Braatz, Fault Detection and Diagnosis in Industrial Systems, Springer Verlag, 2001.
[16] S. Joe Qin, Statistical process monitoring: basics and beyond, J. Chemometr. 17 (2003) 480–502.
[17] V. Venkatasubramanian, R. Rengaswamy, S.N. Kavuri, K. Yin, A review of process fault detection and diagnosis: part III: process history based methods, Comput. Chem. Eng. 27 (2003) 327–346.
[18] P. Cui, J. Li, G. Wang, Improved kernel principal component analysis for fault detection, Expert Syst. Appl. 34 (2008) 1210–1219.
[19] Z. Ge, C. Yang, Z. Song, Improved kernel PCA-based monitoring approach for nonlinear processes, Chem. Eng. Sci. 64 (2009) 2245–2255.
[20] M.A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, AICHE J. 37 (1991) 233–243.
[21] D. Dong, T.J. McAvoy, Nonlinear principal component analysis—based on principal curves and neural networks, Comput. Chem. Eng. 20 (1996) 65–78.
[22] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1998) 1299–1319.
[23] S.W. Choi, I.-B. Lee, Nonlinear dynamic process monitoring based on dynamic kernel PCA, Chem. Eng. Sci. 59 (2004) 5897–5908.
[24] S.W. Choi, J. Morris, I.-B. Lee, Nonlinear multiscale modelling for fault detection and identification, Chem. Eng. Sci. 63 (2008) 2252–2266.
[25] Z. Ge, Z. Song, Kernel generalization of PPCA for nonlinear probabilistic monitoring, Ind. Eng. Chem. Res. 49 (2010) 11832–11836.
[26] Y. Zhang, S. Li, Y. Teng, Dynamic processes monitoring using recursive kernel principal component analysis, Chem. Eng. Sci. 72 (2012) 78–86.
[27] M. Rashid, J. Yu, Nonlinear and non-Gaussian dynamic batch process monitoring using a new multiway kernel independent component analysis and multidimensional mutual information based dissimilarity method, industrial & engineering chemistry research, Ind. Eng. Chem. Res. 51 (2012) 10910–10920.
[28] Q. Chen, U. Kruger, M. Meronk, A. Leung, Synthesis of T^2 and Q statistics for process monitoring, Control Eng. Pract. 12 (2004) 745–755.
[29] H. Wang, Z. Song, P. Li, Fault detection behavior and performance analysis of principal component analysis based process monitoring methods, Ind. Eng. Chem. Res. 41 (2002) 2455–2464.
[30] Q. Jiang, X. Yan, Chemical processes monitoring based on weighted principal component analysis and its application, Chemom. Intell. Lab. Syst. 119 (2012) 11–20.
[31] S. Wold, Exponentially weighted moving principal components analysis and projections to latent structures, Chemom. Intell. Lab. Syst. 23 (1994) 149–161.
[32] X.B. He, Y.P. Yang, Y.H. Yang, Fault diagnosis based on variable-weighted kernel Fisher discriminant analysis, Chemom. Intell. Lab. Syst. 93 (2008) 27–33.
[33] D.L. Ferreira, S. Kittiwachana, L.A. Fido, D.R. Thompson, R.E. Escott, R.G. Brereton, Multilevel simultaneous component analysis for fault detection in multicampaign process monitoring: application to on-line high performance liquid chromatography of a continuous process, Analyst 134 (2009) 1571–1585.
[34] M. Rosenblatt, Curve estimates, Ann. Math. Stat. (1971) 1815–1842.
[35] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Stat. 33 (1962) 1065–1076.
[36] A.R. Webb, K.D. Copsey, G. Cawley, Statistical Pattern Recognition, Wiley, 2011.
[37] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC, 1986.
[38] R.A. Johnson, Applied Multivariate Statistical Analysis, Prentice-Hall, New Jersey, 2007.
[39] S. Yoon, J.F. MacGregor, Fault diagnosis with multivariate statistical models part I: using steady state fault signatures, J. Process. Control. 11 (2001) 387–400.
[40] H.H. Yue, S.J. Qin, Reconstruction-based fault identification using a combined index, Ind. Eng. Chem. Res. 40 (2001) 4403–4414.
[41] S.W. Choi, C.K. Yoo, I.-B. Lee, Overall statistical monitoring of static and dynamic patterns, Ind. Eng. Chem. Res. 42 (2003) 108–117.
[42] J.J. Downs, E.F. Vogel, A plant-wide industrial process control problem, Comput. Chem. Eng. 17 (1993) 245–255.
[43] T. McAvoy, N. Ye, Base control for the Tennessee Eastman problem, Comput. Chem. Eng. 18 (1994) 383–413.
[44] P.R. Lyman, C. Georgakis, Plant-wide control of the Tennessee Eastman problem, Comput. Chem. Eng. 19 (1995) 321–331.
[45] J.M. Lee, S.J. Qin, I.B. Lee, Fault detection of non-linear processes using kernel independent component analysis, Can. J. Chem. Eng. 85 (2007) 526–536.