

Statistical Monitoring of Chemical Processes Based on Sensitive Kernel Principal Components^{*}

JIANG Qingchao (姜庆超) and YAN Xuefeng (颜学峰)^{**}

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

Abstract The kernel principal component analysis (KPCA) method employs the first several kernel principal components (KPCs), which indicate the most variance information of normal observations for process monitoring, but may not reflect the fault information. In this study, sensitive kernel principal component analysis (SKPCA) is proposed to improve process monitoring performance, *i.e.*, to deal with the discordance of T^2 statistic and squared prediction error δ_{SPE} statistic and reduce missed detection rates. T^2 statistic can be used to measure the variation directly along each KPC and analyze the detection performance as well as capture the most useful information in a process. With the calculation of the change rate of T^2 statistic along each KPC, SKPCA selects the sensitive kernel principal components for process monitoring. A simulated simple system and Tennessee Eastman process are employed to demonstrate the efficiency of SKPCA on online monitoring. The results indicate that the monitoring performance is improved significantly.

Keywords statistical process monitoring, kernel principal component analysis, sensitive kernel principal component, Tennessee Eastman process

1 INTRODUCTION

As production equipment in modern chemical industry becomes larger, speed-higher, more complex and more intelligent, fault detection and diagnosis in chemical processes are more important for safety and improvement of productivity. Recently, multivariate statistical process monitoring (MSPM) methods, which identify linear relationships between the recorded process variables and generate reduced sets of scores, have great progress because of the development of equipment for measurement, data storage and computation [1–5]. Principal component analysis (PCA) is the most widely used method because it can effectively deal with high-dimensional, noisy and highly correlated data by projecting the data onto a lower-dimensional space that contains most variance information of data [6–8].

To deal with the problem from nonlinear data, Kramer developed a nonlinear PCA method based on auto-associative neural networks [9], but it is difficult for the network to be trained because it has five layers and to determine the number of nodes in each layer. Dong and McAvoy [6] developed a nonlinear PCA approach based on principal curves and neural networks, with the assumption that the nonlinear function can be approximated by a linear combination of several univariate functions so that it can be expressed as a sum of functions of individual variables. Such mappings can be made only for a limited class of nonlinear models and the application of the principal curve algorithm is restricted to the identification of structures that exhibit additive-type behavior [10, 11]. Many extensional methods of PCA, such as kernel PCA

(KPCA), dynamic PCA (DPCA) [12–15], probabilistic PCA (PPCA) [16, 17] and multiway PCA (MPCA) [18, 19], have been proposed to improve process monitoring performance, among which KPCA, as a new nonlinear PCA technique, progresses fast as a promising method for tackling nonlinear systems [20–22]. KPCA first maps the input space into a feature space *via* nonlinear mapping and then computes the principal components (PCs) in the high-dimensional feature space. Compared to other nonlinear methods, the main advantage of KPCA is that it does not involve nonlinear optimization.

In KPCA monitoring, the first several kernel principal components (KPCs) corresponding to the largest eigenvalues based on normal observations are selected, without considering the information of process fault. However, the number of components in high-dimensional space is large, and the sensitivities of principal components to faults are not the same and nor in the order of eigenvalue decrease. Process monitoring based on KPCA has a problem, *i.e.*, T^2 statistic and squared prediction error (δ_{SPE}) statistic do not give the same result and miss detection when a fault occurs, which are mostly due to missing fault information in process data compressing, so only the first several KPCs are employed for process monitoring. Hence, sensitive kernel principal components that represent the character information of the fault are expected to be found and can be used to monitor the process for good monitoring performance.

In this article, a novel fault detection method based on sensitive kernel principal component analysis (SKPCA) is proposed to improve process monitoring performance. The definition of sensitive kernel

Received 2011-11-29, accepted 2012-08-30.

^{*} Supported by the 973 project of China (2013CB733600), the National Natural Science Foundation (21176073), the Doctoral Fund of Ministry of Education (20090074110005), the New Century Excellent Talents in University (NCET-09-0346), “Shu Guang” project (09SG29) and the Fundamental Research Funds for the Central Universities.

^{**} To whom correspondence should be addressed. E-mail: xfyang@ecust.edu.cn

principal component (SKPC) is introduced and the change rate of T^2 statistic is defined to determine SKPCs of some fault due to the ability of T^2 statistic to measure the variation directly along each kernel principal component. These SKPCs contain large amount of information for fault detection and diagnosis, which is illustrated by observing the monitoring behavior of T^2 statistics along each KPC on a simulated simple process. The case studies on the simulated process and Tennessee Eastman (TE) process demonstrate the performance of the SKPCA model on online monitoring.

2 PROCESS MONITORING BASED ON KERNEL PCA

2.1 Kernel principal component analysis

In KPCA, observations are nonlinearly mapped into a high-dimensional space, referred to as feature space F . In the feature space, the linear PCA is employed to extract the nonlinear correlation relationship between the variables. According to Cover's theorem, the nonlinear data structure in the input space is more likely to be linear after high-dimensional nonlinear mapping [23]. KPCA finds a computationally tractable solution through a simple kernel function that intrinsically constructs a nonlinear mapping from the input space to the feature space [24, 25].

Let the normalized training set be $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbf{R}^m$, with N observations consisting of m measured process variables. The feature space is constructed by using a nonlinear mapping: $\mathbf{R}^m \xrightarrow{\Phi(\cdot)} \mathbf{C}^F$, where $\Phi(\cdot)$ is a nonlinear mapping function. Similar to PCA, the covariance matrix in the feature space F can be calculated as

$$\mathbf{C}^F = \frac{1}{N} \sum_{j=1}^N \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T \quad (1)$$

in which $\sum_{j=1}^N \Phi(\mathbf{x}_j) = 0$ is assumed. The kernel principal component can be obtained by the eigenvalue problem below

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

where λ and $\mathbf{v}$ denote eigenvalue and eigenvector of covariance matrix $\mathbf{C}^F$, respectively, and $\langle \mathbf{x}, \mathbf{y} \rangle$ 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 $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_N)$. Then $\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v}$ is equivalent to

$$\lambda \langle \Phi(\mathbf{x}_k), \mathbf{v} \rangle = \langle \Phi(\mathbf{x}_k), \mathbf{C}^F \mathbf{v} \rangle \quad (k=1, \dots, N) \quad (3)$$

There exist coefficients α_i ($i=1, \dots, N$), so that

$$\mathbf{v} = \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_i) \quad (4)$$

Combining Eqs. (3) and (4), we obtain

$$\begin{aligned} \lambda \sum_{i=1}^N \alpha_i \langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_i) \rangle = \\ \frac{1}{N} \sum_{i=1}^N \alpha_i \left\langle \Phi(\mathbf{x}_k), \sum_{j=1}^N \Phi(\mathbf{x}_j) \right\rangle \langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}_i) \rangle \\ (k=1, \dots, N) \end{aligned} \quad (5)$$

Note that the eigenvalue problem in Eq. (5) only involves dot products of mapped shape vectors in the feature space. We define an $N \times N$ matrix $\mathbf{K}$ by $[\mathbf{K}]_{ij} = K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$. Then the left-hand side of Eq. (5) can be expressed as

$$\lambda \sum_{i=1}^N \alpha_i \langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_i) \rangle = \lambda \sum_{i=1}^N \alpha_i K_{ki} \quad (k=1, \dots, N) \quad (6)$$

and the right-hand side of Eq. (5) can be expressed as

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \alpha_i \left\langle \Phi(\mathbf{x}_k), \sum_{j=1}^N \Phi(\mathbf{x}_j) \right\rangle \langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}_i) \rangle = \\ \frac{1}{N} \sum_{i=1}^N \alpha_i \sum_{j=1}^N K_{kj} K_{ji} \\ (k=1, \dots, N) \end{aligned} \quad (7)$$

Equation (7) becomes $(1/N) \mathbf{K}^2 \boldsymbol{\alpha}$. Combining Eqs. (6) and (7), we obtain

$$\lambda N \mathbf{K} \boldsymbol{\alpha} = \mathbf{K}^2 \boldsymbol{\alpha} \quad (8)$$

where $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^T$. To find solutions of Eq. (8), we solve the eigenvalue problem

$$N \lambda \boldsymbol{\alpha} = \mathbf{K} \boldsymbol{\alpha} \quad (9)$$

for nonzero eigenvalues. Now, performing PCA in F is equivalent to resolving the eigenvalue-problem of Eq. (9). This yields eigenvectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_N$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$. The dimensionality of the problem can be reduced by retaining only the first p eigenvectors using the cumulative contribution rate r_{CCR} method.

$$r_{CCR} = \sum_{i=1}^p \lambda_i / \sum_{i=1}^N \lambda_i \times 100\% \quad (10)$$

We normalize $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_p$ by requiring that the corresponding vectors in F be normalized, i.e.

$$\langle \mathbf{v}_k, \mathbf{v}_k \rangle = 1 \quad (k=1, \dots, p) \quad (11)$$

Using $\mathbf{v}_k = \sum_{i=1}^N \alpha_i^k \Phi(\mathbf{x}_i)$, Eq. (11) leads to

$$\begin{aligned}
 1 &= \left\langle \sum_{i=1}^N \alpha_i^k \Phi(\mathbf{x}_i), \sum_{j=1}^N \alpha_j^k \Phi(\mathbf{x}_j) \right\rangle \\
 &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i^k \alpha_j^k \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \\
 &= \langle \boldsymbol{\alpha}_k, \mathbf{K} \boldsymbol{\alpha}_k \rangle = \lambda_k \langle \boldsymbol{\alpha}_k, \boldsymbol{\alpha}_k \rangle
 \end{aligned}$$

The principal component t of a test vector $\mathbf{x}$ are then extracted by projecting $\Phi(\mathbf{x})$ onto eigenvectors $\mathbf{v}_k$ in F ,

$$\begin{aligned}
 t_k &= \langle \mathbf{v}_k, \Phi(\mathbf{x}) \rangle = \sum_{i=1}^N \alpha_i^k \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle \\
 (k &= 1, \dots, p)
 \end{aligned} \quad (12)$$

To solve the eigenvalue problem of Eq. (9) and to project from the input space into the KPCA space using Eq. (12), one can avoid performing the nonlinear mappings and computing the dot products in the feature space by introducing a kernel function of form $K(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$.

There exist a number of kernel functions and in this work the radial basis kernel function is used, which is

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{c}\right) \quad (13)$$

where c is specified a priori by the user.

Before applying KPCA, mean centering in the high-dimensional space should be performed. This can be done by substituting the kernel matrix $\mathbf{K}$ with

$$\bar{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K} \mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N \quad (14)$$

$$\text{where } \mathbf{1}_N = \frac{1}{N} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \in \mathbf{R}^{N \times N}.$$

2.2 T^2 statistic and δ_{SPE} statistic

T^2 statistic and δ_{SPE} statistic are constructed and monitored in KPCA based on the assumption that training data are multivariate normal in the feature space [10, 24]. T^2 statistic is the sum of the normalized squared scores and is defined as

$$T^2 = [t_1, \dots, t_p] \mathbf{A}^{-1} [t_1, \dots, t_p]^T \quad (15)$$

where t_k is obtained from Eq. (12) and $\mathbf{A}^{-1}$ is the diagonal matrix of the inverse of the eigenvalues with the retained PCs. The confidence limit for T^2 is obtained using the F -distribution

$$T_{p,N,\gamma}^2 \sim \frac{p(N-1)}{N-p} F_{p,N-p,\gamma} \quad (16)$$

where N is the number of samples in the model, p is the number of PCs, and γ is the level of significance.

In KPCA model, the δ_{SPE} statistic is obtained

using the equations

$$\begin{aligned}
 \delta_{\text{SPE}} &= \|\Phi(\mathbf{x}) - \Phi_p(\mathbf{x})\| = \|\bar{\Phi}(\mathbf{x}) - \bar{\Phi}_p(\mathbf{x})\| \\
 &= \sum_{j=1}^n t_j^2 - \sum_{j=1}^p t_j^2
 \end{aligned} \quad (17)$$

where n is the number of nonzero eigenvalues generated from Eq. (9) among the total N eigenvalues. The confidence limit for the δ_{SPE} can be computed from its approximate distribution

$$\delta_{\text{SPE},\gamma} \sim g \chi_h^2 \quad (18)$$

where g is a weighting parameter to account for the magnitude of the δ_{SPE} , χ^2 is the chi-square distribution and h accounts for the degree of freedom. More details on the two statistic parameters are represented in [10, 24].

2.3 Process monitoring based on KPCA

The following simulations are in Matlab 7.12.0 (2011a) environment. Consider the following simple system with three variables but only one factor, originally suggested by Dong and McAvoy [6, 24].

$$x_1 = t + e1 \quad (19)$$

$$x_2 = t^2 - 3t + e2 \quad (20)$$

$$x_3 = -t^3 + 3t^2 + e3 \quad (21)$$

where $e1$, $e2$ and $e3$ are independent noise variables $N(0, 0.01)$, and $t \in [0.01, 2]$. 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 disturbances are applied separately during generation of the test data sets.

Disturbance 1: a step change of x_2 by -0.4 is introduced starting from sample 101.

Disturbance 2: x_1 is linearly increased from sample 101 to 270 by adding $0.01 \times (s - 100)$ to the x_1 value of each sample in this range, where s is the sample number.

Monitoring performance of KPCA is shown in Fig. 1, in which the first 3 kernel principal components occupying 85% cumulative contribution rate are retained. For Disturbance 1 in this simple system, the disturbance can be detected successfully by δ_{SPE} statistic, while T^2 statistic does not behave very well. For Disturbance 2, both T^2 statistic and δ_{SPE} statistic detect the disturbance successfully.

In the simple system, there are 3 variables and the number of kernel principal components may be arbitrary when mapped into feature space. In order to illustrate the detecting behavior of T^2 statistic, the first 20 kernel principal components are taken into consideration. The T^2 statistic of each kernel principal component is calculated, and there are 20 subfigures for the process monitoring. For Disturbance 1, the results

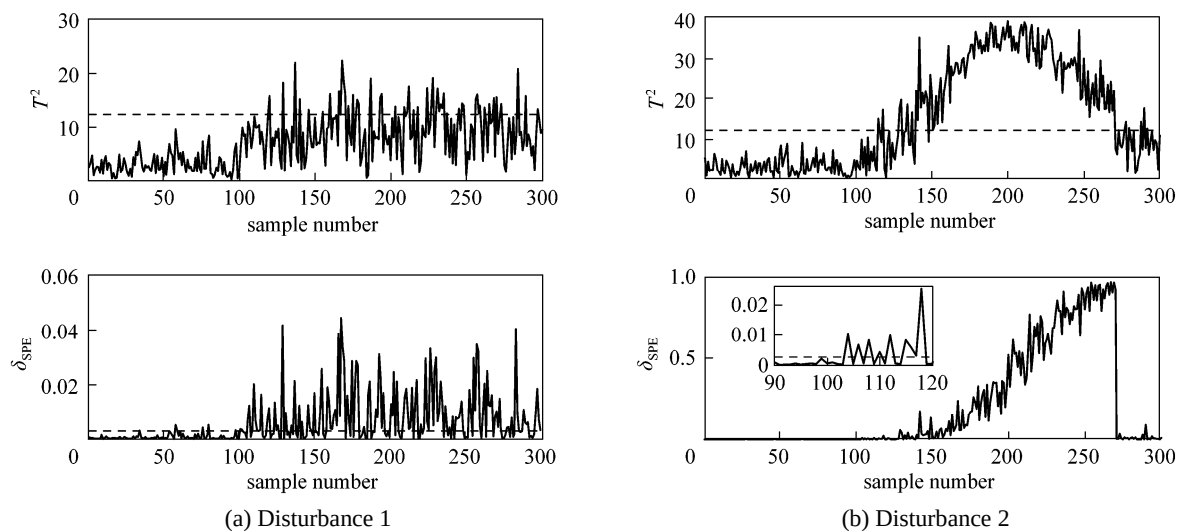


Figure 1 Online monitoring of Disturbance 1 (a) and Disturbance 2 (b) using KPCA model

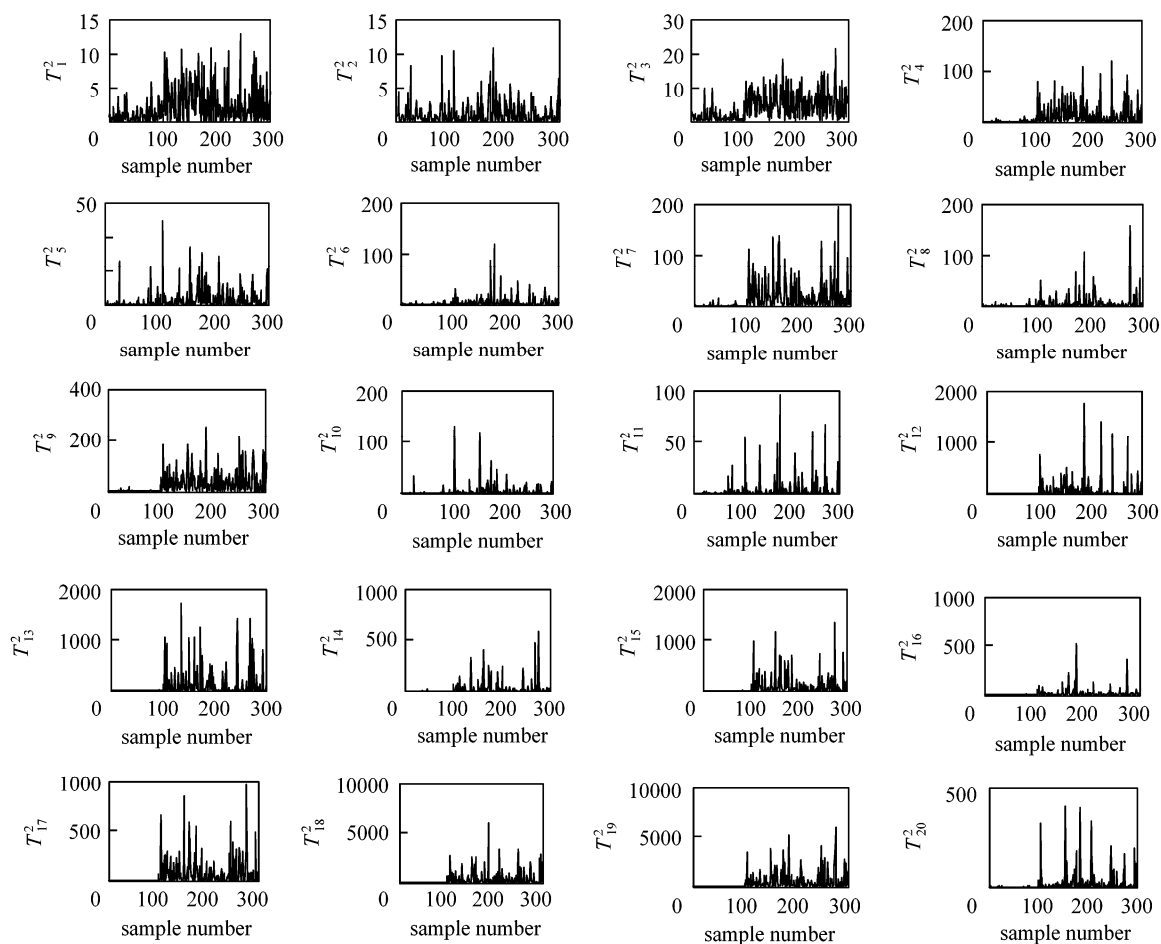


Figure 2 T_k^2 of each kernel principal component

are shown in Fig. 2, in which 20 subfigures are in the order of eigenvalues decrease and $T_k^2 = t_k \lambda_k^{-1} t_k^T$ represents T^2 statistic of the k th kernel principal component, where $m=1,2,\dots,20$. In Fig. 2, the first 3 kernel principal components corresponding to the

largest 3 eigenvalues do not show the largest variation when Disturbance 1 occurs using T_k^2 statistic, and T_k^2 of the 9th, 16th, 17th and 19th kernel principal components change remarkably when disturbance 1 comes. The quantitative analysis of T_k^2 change rate will be

executed in Section 3, since the kernel principal component directions are defined as those with maximum variance of normal operating process sample, without considering the information of fault samples. The T^2 statistic directly measures the variation along each of the loading vectors, *i.e.*, the directions of each kernel principal component. The directions in which the fault samples show larger variation using T^2 statistic are not always the same as the loading vectors corresponding to the larger eigenvalues. Hence T^2 statistic of the kernel principal components chosen according to the larger eigenvalues may miss the important information of fault, and lead to missed detection problem.

3 PROCESS MONITORING BASED ON SKPCA

3.1 Sensitive kernel principal components

The analysis in Section 2 reveals that the T^2 statistic in some directions of kernel principal components changes more remarkably than others when a fault occurs. Namely, in these directions, the fault sample data show large variation and the fault can be detected easily when using T^2 statistic corresponding to these kernel principal components. To describe the variation in the direction of the k th kernel principal component, the change rate of T_k^2 at the i th sample point is defined as

$$R_{T_{k,i}^2} = T_{k,i}^2 / \overline{T_k^2} \quad (22)$$

where $T_{k,i}^2$ is the T^2 statistic of the k th kernel principal component on the i th measurement data, and $\overline{T_k^2}$ is the mean value of T_k^2 based on a normal sample, *i.e.*

$$\overline{T_k^2} = \frac{1}{N} \sum_{i=1}^N T_{k,i}^2.$$

The threshold of $R_{T_{k,i}^2}$ could be determined using kernel density estimation based on normal operating sample data since $R_{T_{k,i}^2}$ over some period does not follow a Gaussian distribution and the threshold can not be determined directly from a particular approximate distribution [26]. Kernel density estimation, in which there is no formal structure for the density prescribed, is a nonparametric density estimation method. A univariate kernel estimator with kernel K is defined as

$$\hat{f}(x) = \frac{1}{cN} \sum_{i=1}^N K\left\{\frac{x-x_i}{c}\right\} \quad (23)$$

where x is the data point under consideration, x_i is an observation value from the data set, c is the window width (also known as the smoothing parameter), N is the number of observations, and K is the kernel function. The kernel function K determines the shape of the smooth curve and satisfies the condition

$$\int_{-\infty}^{+\infty} K(x)dx = 1 \quad (24)$$

There are a number of possible kernel functions and the Gaussian kernel function is the most commonly used. In our trial, Gaussian kernel function is used. The threshold of $R_{T_{k,i}^2}$ is determined in following steps.

(1) Calculate $R_{T_{k,i}^2}$ of the m th kernel principal component on the i th sample point based on normal operating sample data.

(2) The univariate kernel density estimator is used to estimate the density function of $R_{T_k^2}$ on the normal operation condition.

(3) Determine the control limit of each component, which occupies the 90% area of density function and is denoted by l_{CL}^k , namely $\hat{f}(l_{CL}^k) \geq 90\%$ (the control limit of each component is an empirical value and it could be modified according to the practical situation. The advised range of the control limit is 85% and 95%, and 90% is used in this work).

(4) The threshold of $R_{T_{k,i}^2}$ is the maximum one among l_{CL}^k and represented by l_{CL} , *i.e.* $l_{CL} = \max\{l_{CL}^k, k=1, 2, \dots, p\}$, in which p is the number of principal components.

The kernel principal components, in which directions the T^2 statistic has larger $R_{T_k^2}$ when a fault occurs, could be considered as sensitive kernel principal components to this fault. The m th kernel principal component is regarded as sensitive principal component on the condition that

$$\overline{R_{T_{k,a}^2}} = \frac{1}{n} \sum_{i=a-n+1}^a R_{T_{k,i}^2} \geq l_{CL} \quad (25)$$

where a is the current sample point, n is an empirical value and usually decided according to the sample time. The recommend value of n in this paper is 6, which is the same as the number of points in fault distinguish rule in Shewhart's theory [5, 27]. Using the mean value $\overline{R_{T_k^2}}$ can effectively decrease the influence of process noise. The kernel principal component corresponding to the largest value $\overline{R_{T_k^2}}^{\max}$ among $\overline{R_{T_k^2}}$ could be regarded as the most sensitive kernel principal component to the fault, where

$$\overline{R_{T_k^2}}^{\max} = \max\{\overline{R_{T_k^2}}, n=1, 2, \dots, p\} \quad (26)$$

3.2 Process monitoring using sensitive kernel principal components

Based on the analysis above, it is obvious that to

monitor the fault is to monitor the T^2 statistic with the largest change rate, *i.e.* $\overline{R_{T_k^2}^{\max}}$. If $\overline{R_{T_k^2}^{\max}} < l_{CL}$, the process is regarded as normal. If $\overline{R_{T_k^2}^{\max}} \geq l_{CL}$, there may be a fault in process and further analysis is needed. Here a new modeling way called process monitoring based on sensitive kernel principal components is proposed and the calculation steps are as follows.

Off line modeling:

(1) Divide the collected normal operating sample data into two sets, one as training data and another for testing. Let the training data be $X \in R^{N \times m}$, where N is the number of observations consisting of m measurement process variables.

(2) Normalize the data using the mean and standard deviation of each variable. Compute the kernel matrix $K \in R^{N \times N}$ by $[K]_{ij} = \langle \Phi(x_i), \Phi(x_j) \rangle = [K(x_i, x_j)]$.

(3) Carry out centering in the feature space for $\sum_{k=1}^N \overline{\Phi}(x_k) = 0$, $\overline{K} = K - \mathbf{1}_N K - K \mathbf{1}_N + \mathbf{1}_N K \mathbf{1}_N$.

(4) Solve the eigenvalue problem $N\lambda\alpha = \overline{K}\alpha$ and normalize α_k such that $\langle \alpha_k, \alpha_k \rangle = 1/\lambda_k$.

(5) For normal operating data x , extract a nonlinear component via $t_k = \langle v_k, \overline{\Phi}(x) \rangle = \sum_{i=1}^N \alpha_i^k \langle \overline{\Phi}(x_i), \overline{\Phi}(x) \rangle = \sum_{i=1}^N \alpha_i^k \overline{K}(x_i, x)$.

(6) Based on testing data, examine the T^2 statistic of each component T_k^2 , and calculate $R_{T_{k,j}^2}$ of the normal operation process testing data. Because the dimension in KPCA may be arbitrarily high, a proper number of kernel principal components should be examined (usually, the advised control limit of r_{CCR} is between 99% and 99.5%, and only abandon those corresponding to some smallest eigenvalues. Here 99% is used, namely $r_{CCR} \geq 99\%$).

(7) Use kernel density estimation to determine the threshold l_{CL}^k of each component and the control limit l_{CL} of a normal process. Save l_{CL}^k and l_{CL} .

Online monitoring:

(1) Normalize the current sample data using mean values and variance of the training data.

(2) Calculate $T_{k,i}^2$, $R_{T_{k,i}^2}$, $\overline{R_{T_k^2}^{\max}}$ and $\overline{R_{T_k^2}^{\max}}$ of the current data.

(3) Determine if the current $\overline{R_{T_k^2}^{\max}}$ runs out of the control limit l_{CL} or not. If not, repeat Steps (1) and (2) to monitor the next sample point. If $\overline{R_{T_k^2}^{\max}} \geq l_{CL}$,

there may be a fault and go to the following steps.

(4) Select the sensitive kernel principal components according to $\overline{R_{T_k^2}} = \frac{1}{n} \sum_{i=a-n+1}^a R_{T_{k,i}^2} \geq l_{CL}$. Let the number of sensitive kernel principal components be q and these q sensitive kernel principal components are denoted as k_1 th, k_2 th, $\dots$, k_q th kernel principal components.

(5) Use the selected q sensitive kernel principal components to monitor the process, and determine the control limits of T^2 with q kernel principal components according to formula $\delta_T^2 = \frac{q(N-1)}{N-q} F_{q, N-q, \gamma}$.

(6) Calculate T^2 statistic of the selected k_1 sensitive kernel principal components, which is represented by $T^2(q)$. If $T^2(q) < \delta_T^2$, go to Step (1). If $T^2(q) \geq \delta_T^2$, there is a fault in process.

3.3 Simulations and application

3.3.1 Application in the simple system

Process monitoring approach based on sensitive kernel principal component is applied in the simple system with three variables. Fig. 3 shows $R_{T_{1,j}^2}$ and cumulative probability density of $R_{T_{1,j}^2}$ for the first kernel principal component as example, the control limits l_{CL}^k of 20 kernel principal components and the threshold l_{CL} of $R_{T_{1,j}^2}$ on the normal operating condition. The control limits l_{CL}^k of 20 kernel principal components are corresponding to 99% area of their density functions, and the threshold l_{CL} of $R_{T_{1,j}^2}$ is 31.74 for this simple process.

In the simple system, Disturbance 1 is employed to demonstrate the online monitoring performance of SKPCA model to compare with KPCA model. The eigenvalues of normal sample data and $\overline{R_{T_k^2}}$ of each kernel principal components are shown in Fig. 4 when Disturbance 1 occurs. The first 3 kernel principal components correspond to larger eigenvalues while the T_k^2 statistic of the 9th, 16th, 17th and 19th kernel principal component change more remarkably when Fault 2 occurs.

The monitoring performance of Disturbance 1 in the simple system using SKPCA is shown in Fig. 5. The first subfigure describes $\overline{R_{T_k^2}^{\max}}$, designated as the maximum value of change rate of T_k^2 ($R_{T_k^2}^{\max}$) for convenient. The last two subfigures are the monitoring

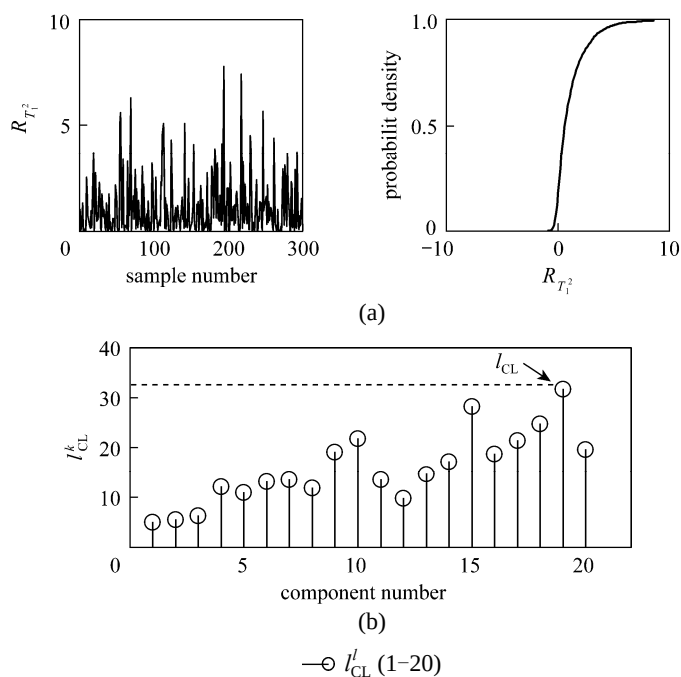


Figure 3 Cumulative probability density of $R_{T_{ii}}^2$ (a), l_{CL}^k and l_{CL} of the simple system (b)

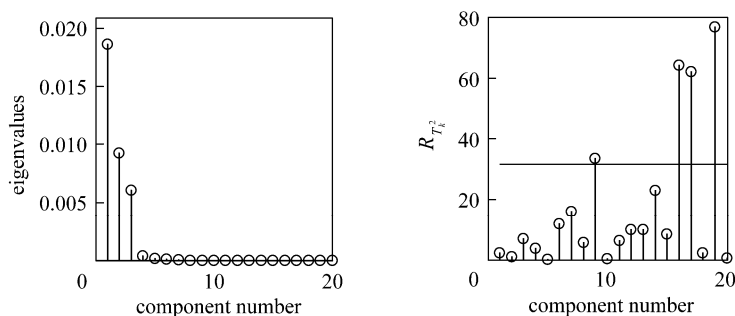


Figure 4 Eigenvalues of normal sample and $\overline{R_{T_k}^2}$ of Disturbance 2

performance using SKPCA. SKPC T^2 and SKPC δ_{SPE} are T^2 statistic and δ_{SPE} statistic calculated using sensitive kernel principal components according to Eqs. (15)–(18) with q sensitive kernel principal components presented in Section 2.2. Fig. 5 shows that Disturbances 1 and 2 can be detected successfully and timely by T^2 statistic using SKPCA model. Compared with the results in Fig. 1, it is clear that SKPC T^2 detects Disturbances 1 and 2 much earlier than KPCA.

3.3.2 Application in TE process

TE process is a benchmark problem in chemical process engineering. Downs and Vogel presented this particular process at an AIChE meeting in 1990s as a plant-wide control problem [28]. This simulator of Tennessee Eastman process consists of five major unit operations: a reactor, a product condenser, a vapor-liquid separator, a recycle compressor and a product stripper. Two products are produced by two simultaneous gas-liquid exothermic reactions, and a byproduct

is from two additional exothermic reactions. The process has 12 manipulated variables, 22 continuous process measurements and 19 compositions. The simulator can generate 21 types of different faults. All the process measurements include Gaussian noise. Once the fault enters the process, it affects almost all state variables in the process [29, 30]. The control scheme for the Tennessee Eastman process is shown in Fig. 6.

Monitoring performance of Faults 11 and 14 in TE process based on sensitive kernel principal components are shown in Figs. 7 and 8. In TE process, Fault 11 involves a step change in inlet temperature of reactor cooling water and Fault 14 is the sticking of reactor cooling water valve. Both faults enter the process at the 161th sample point. Fig. 7 shows the change rate of T_k^2 of the two faults at the 163th sample point, the 4th, 35th, 39th, 43th, 45th and 51th components are the sensitive kernel principal components of Fault 11, and the 32th, 35th, 39th, 43th, 44th and 58th components are the sensitive kernel principal components of Fault 14.

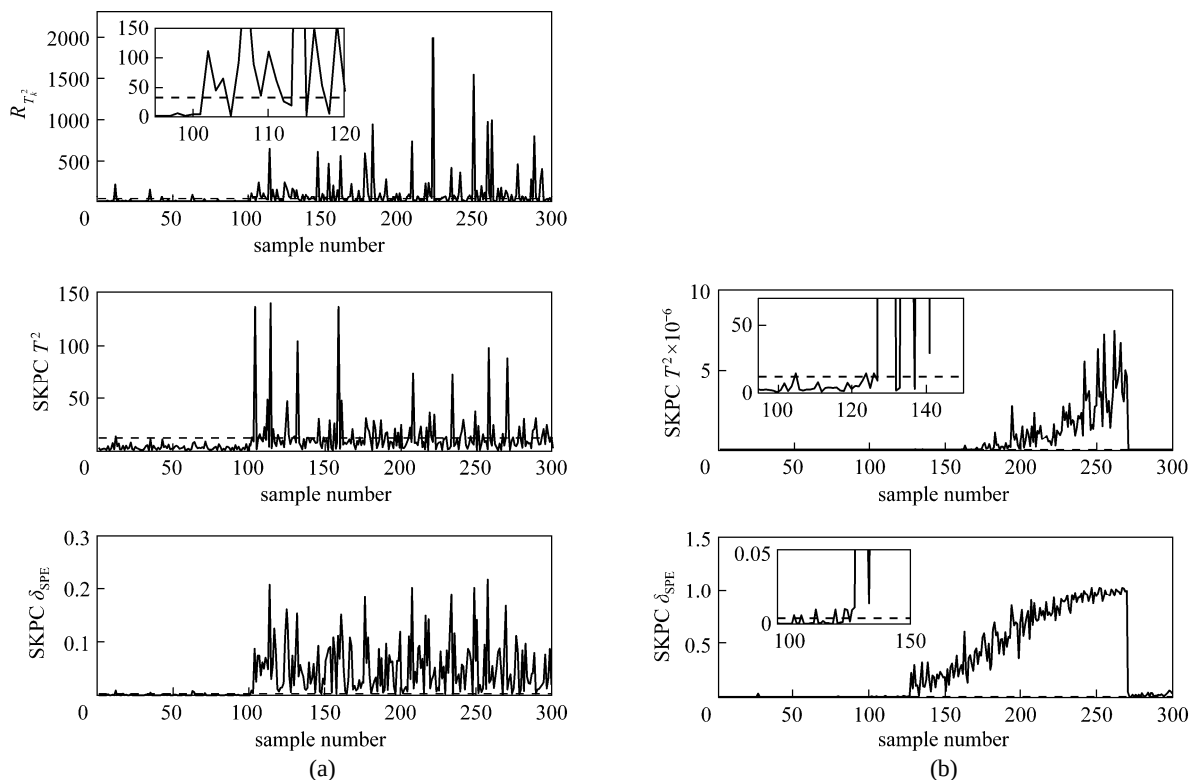


Figure 5 Monitoring performance of Disturbances 1 (a) and 2 (b) using SKPCA

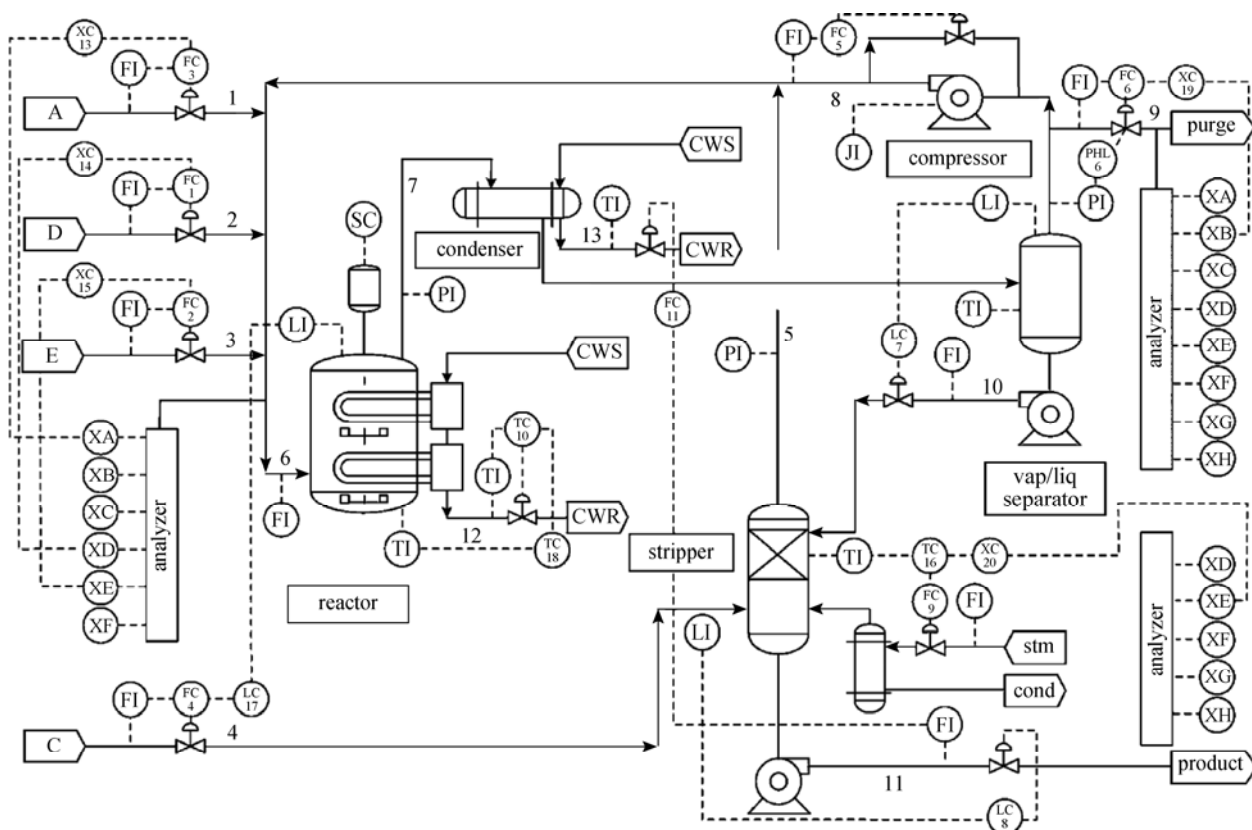


Figure 6 Control scheme for the Tennessee Eastman process
CWS: cooling water suck; CWR: cooling water return; A-E: components

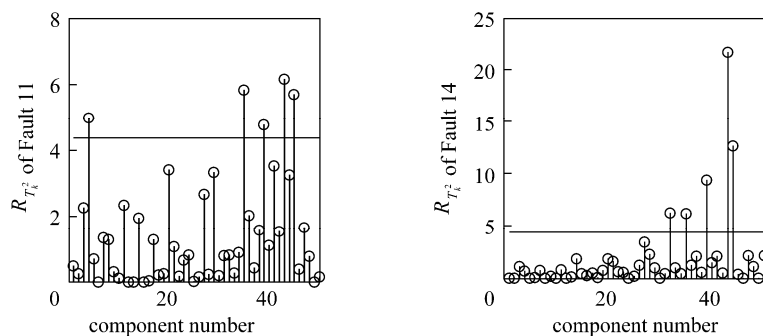


Figure 7 Sensitive kernel principal components of Faults 11 and 14

○ $R_{T_k^2}$; — l_{CL}

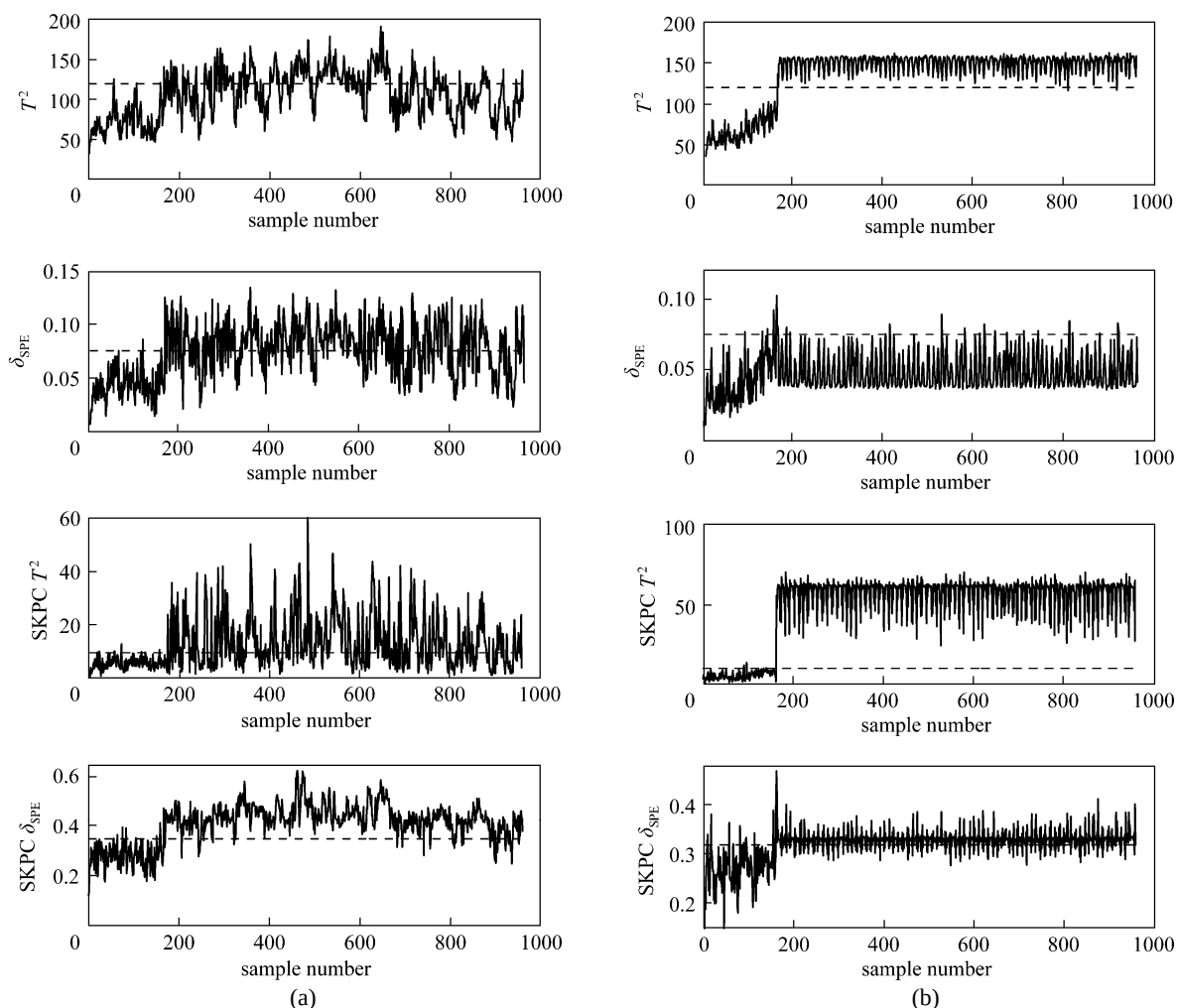


Figure 8 Monitoring performance of Faults 11 (a) and 14 (b) using KPCA and SKPCA models

Figure 8 presents the monitoring performance based on KPCA and SKPCA models and reveals that the T^2 statistic of the sensitive kernel principal components can detect the faults successfully and timely. When Fault 11 occurs, T^2 with SKPC presents a large jump and is more than the threshold. The result indicates that there is a fault in process. When Fault 14 occurs,

both SKPC T^2 statistic and SKPC δ_{SPE} statistic can detect the fault successfully, showing better monitoring performance than that in KPCA model.

T^2 statistic is an important statistic when making a judgment for the process state. Process monitoring based on SKPCA takes the information of fault into account timely and can catch the largest variation in

process, improving the fault detection performance of T^2 statistic significantly. Once a fault is detected, the next step is to determine the cause of the out-of-control status. The task of fault diagnosis is rather challenging when the number of process variables is large and the process is highly integrated. The sensitive kernel principal components contain a large amount of information of the coming fault, providing a novel approach for fault diagnosis, though more work is needed.

4 CONCLUSIONS

Since KPCA model can not detect some faults efficiently, the behavior of T^2 statistic and the monitoring performance are analyzed to find the cause. In order to improve the online monitoring performance, process monitoring based on sensitive kernel principal components is proposed. SKPCA examines T^2 statistic of each kernel principal component and searches the most sensitive kernel principal components of fault in feature space, which is applied to determine whether there is a fault in process or not. The novel modeling method is applied to online monitoring of a simple process and TE process. Compared with the classical KPCA models, the monitoring performance of SKPCA is improved. Future work will be focused on fault diagnosis based on SKPCA. The sensitive kernel principal components contain more information of some faults, may represent the fault character and provide significant information to facilitate process fault diagnosis.

NOMENCLATURE

C^F	covariance matrix in feature space
c	smoothing parameter
$e1, e2$ and $e3$	independent noise variables
$\hat{f}$	kernel estimator
g	weighting parameter
h	degree of freedom
K	kernel function
K	kernel matrix
l_{CL}	maximum value of l_{CL}^k
l_{CL}^k	control limit of each component of $R_{T_k^2}$
$R_{T_k^2}^{\max}$	maximum value of change rate of T_k^2
m	number of variables
n	number of observations
p	number of principal components retained
q	number of sensitive kernel principal components
$R_{T_{k,i}^2}$	change rate of T_k^2 on the i th sample point
$\overline{R_{T_k^2}}$	mean value of $R_{T_k^2}$
r_{CCR}	cumulative contribution rate
s	sample number in the simple process
T_k^2	T^2 statistic corresponding to the k th component

$\overline{T_k^2}$	mean value of T_k^2 based on a normal sample
t	principal component
v	eigenvector of covariance matrix
α	eigenvector of kernel matrix
γ	level of significance
δ_T^2	control limit of T^2 statistic
A^{-1}	diagonal matrix of the inverse of eigenvalues
λ	eigenvalue of covariance matrix
$\Phi(\bullet)$	nonlinear mapping function
X	process sample set

REFERENCES

- Lee, J., Kang, B., Kang, S.H., "Integrating independent component analysis and local outlier factor for plant-wide process monitoring", *J. Process Control*, **21** (7), 1011–1021 (2011).
- Li, Y.F., Wang, Z.F., Yuan, J.Q., "On-line fault detection using SVM-based dynamic MPLS for batch processes", *Chin. J. Chem. Eng.*, **14** (6), 754–758 (2006).
- Liu, X.Q., Xie, L., Kruger, U., Littler, T., Wang, S.Q., "Statistical-based monitoring of multivariate non-Gaussian systems", *AIChE J.*, **54** (9), 2379–2391 (2008).
- MacGregor, J., Kourti, T., "Statistical process control of multivariate processes", *Control Eng. Pract.*, **3** (3), 403–414 (1995).
- Zhang, G.X., *An Introduction to The New Multivariate Diagnosis Theory with Two Kinds of Quality*, Science Press, Bei Jing, China (2001).
- Dong, D., McAvoy, T.J., "Nonlinear principal component analysis-based on principal curves and neural networks", *Comput. Chem. Eng.*, **20** (1), 65–78 (1996).
- Kresta, J.V., Macgregor, J.F., Marlin, T.E., "Multivariate statistical monitoring of process operating performance", *Can. J. Chem. Eng.*, **69** (1), 35–47 (2009).
- Wachs, A., Lewin, D.R., "Improved PCA methods for process disturbance and failure identification", *AIChE J.*, **45** (8), 1688–1700 (2004).
- Kramer, M.A., "Nonlinear principal component analysis using auto-associative neural networks", *AIChE J.*, **37** (2), 233–243 (1991).
- Cui, P., Li, J., Wang, G., "Improved kernel principal component analysis for fault detection", *Expert Syst. Appl.*, **34** (2), 1210–1219 (2008).
- Jia, F., Martin, E., Morris, A., "Non-linear principal components analysis with application to process fault detection", *Int. J. Sys. Sci.*, **31** (11), 1473–1487 (2000).
- Jia, M., Chu, F., Wang, F., Wang, W., "On-line batch process monitoring using batch dynamic kernel principal component analysis", *Chemom. Intell. Lab. Syst.*, **101** (2), 110–122 (2010).
- Ku, W., Storer, R.H., Georgakis, C., "Disturbance detection and isolation by dynamic principal component analysis", *Chemom. Intell. Lab. Syst.*, **30** (1), 179–196 (1995).
- Li, R., Rong, G., "Fault isolation by partial dynamic principal component analysis in dynamic process", *Chin. J. Chem. Eng.*, **14** (4), 486–493 (2006).
- Xie, L., Zhang, J., Wang, S., "Investigation of dynamic multivariate chemical process monitoring", *Chin. J. Chem. Eng.*, **14** (5), 559–568 (2006).
- Chen, T., Martin, E., Montague, G., "Robust probabilistic PCA with missing data and contribution analysis for outlier detection", *Comput. Stat. Data An.*, **53** (10), 3706–3716 (2009).
- Kim, D., Lee, I.B., "Process monitoring based on probabilistic PCA",

- Chemom. Intell. Lab. Syst.*, **67** (2), 109–123 (2003).
- 18 Lee, J.M., Yoo, C.K., Lee, I.B., “Fault detection of batch processes using multiway kernel principal component analysis”, *Comput. Chem. Eng.*, **28** (9), 1837–1847 (2004).
- 19 Wang, Z.F., Yuan, J.Q., “Online supervision of penicillin cultivations based on rolling MPCA”, *Chin. J. Chem. Eng.*, **15** (1), 92–96 (2007).
- 20 Choi, S.W., Lee, C., Lee, J.M., Park, J.H., Lee, I.B., “Fault detection and identification of nonlinear processes based on kernel PCA”, *Chemom. Intell. Lab. Syst.*, **75** (1), 55–67 (2005).
- 21 Nguyen, V.H., Golival, J.C., “Fault detection based on Kernel principal component analysis”, *Eng. Struct.*, **32** (11), 3683–3691 (2010).
- 22 Schölkopf, B., Smola, A., Müller, K.R., “Nonlinear component analysis as a kernel eigenvalue problem”, *Neural comput.*, **10** (5), 1299–1319 (1998).
- 23 Haykin, S., *Neural Networks*, Practice-Hall Press, New Jersey, U.S.A. (1999).
- 24 Lee, J.M., Yoo, C.K., Choi, S.W., Vanrolleghem, P.A., Lee, I.B., “Nonlinear process monitoring using kernel principal component analysis”, *Chem. Eng. Sci.*, **59** (1), 223–234 (2004).
- 25 Romdhani, S., Gong, S., Psarrou, A., “A multi-view nonlinear active shape model using kernel PCA”, In: *British Machine Vision Conference*, Nottingham, U.K. (1999).
- 26 Webb, A.R., *Statistical Pattern Recognition*, Oxford University Press, New York, U.S.A. (1999).
- 27 Read, C.B., Kotz, S., Johnson, N.L., *Encyclopedia of Statistical Sciences*, Wiley, Germany (1988).
- 28 Downs, J.J., Vogel, E.F., “A plant-wide industrial process control problem”, *Comput. Chem. Eng.*, **17** (3), 245–255 (1993).
- 29 Lyman, P.R., Georgakis, C., “Plant-wide control of the Tennessee Eastman problem”, *Comput. Chem. Eng.*, **19** (3), 321–331 (1995).
- 30 McAvoy, T., Ye, N., “Base control for the Tennessee Eastman problem”, *Comput. Chem. Eng.*, **18** (5), 383–413 (1994).