Fault Detection of Nonlinear Processes Using Fuzzy C-means-based Kernel PCA

Lamiaa M. Elshenawy, and Tarek A. M. Mohamed

***Abstract*—**Nonlinearity in industrial processes such as chemical and biological processes is still a significant problem. Kernel principal component analysis (KPCA) has recently proven to be a powerful tool for monitoring nonlinear processes with numerous mutually correlated measured variables. One of the drawbacks of original KPCA is that computation time increases with the number of samples. In this article, fuzzy C-means clustering technique (FCM) is adopted to reduce the computational complexity of KPCA. The proposed approach (FCM-KPCA) is applied for fault detection of the Tennessee Eastman chemical process. Simulation results show the effectiveness of the proposed approach in terms of low computational cost and low missed detection rate.

***Keywords*—**Fault detection, fuzzy C-means, kernel PCA, nonlinear processes.

I. Introduction

AMONG the existing nonlinear methods, kernel-based techniques have been successfully developed for tackling the nonlinear problem in recent years [1]. They have attracted wide attentions, including support vector machine (SVM) [2], [3], and [4], kernel principal component analysis (KPCA) [5], [6], [7], [8], and [9], kernel partial least squares (PLS) [10], [11], and [12], kernel fisher discriminant analysis (FDA) [13], [14], [15], and [16] and Kernel Independent Component Analysis (KICA) [17], [18], and [19]. The basic idea is that the mapped data are analyzed using conventional linear statistical analysis techniques in high dimensional feature space, which is equivalent to nonlinear analysis in original input space [20].

KPCA has already shown better performance than PCA in several fields. However, KPCA have characteristics that limit its practical applications in process monitoring. One of the shortcomings of KPCA is that in the training phase of KPCA, it requires to store and manipulate the kernel matrix, the size of which is the square of the number of samples. When the samples number becomes large, the calculation of eigenvalues and eigenvectors will be time consuming [21]. In this paper, we seek to improve KPCA for fault detection in the viewpoint of computational cost. A Fuzzy C-means cluster method [22] is adopted to reduce the computational complexity of KPCA when the number of training samples becomes large. FCM method is used to isolate different classes of time-series data in the input space. Then the clustered data instead of whole input data are mapped into feature space via a nonlinear kernel function to build KPCA for fault detection.

In comparison with the KPCA, the proposed Fuzzy C-means based KPCA (FCM-KPCA) approach effectively reduces the computation cost of the KPCA technique. One of the main objectives of a fault detection statistic is to be sensitive to all possible faults of the process. The sensitivity of the statistics is quantified by calculating the missed detection rates for the faults of the monitored process. In this context, the proposed approach as a nonlinear monitoring model is compared with the PCA in terms of sensitivity. The utility of the proposed monitoring approach is tested using a Tennessee Eastman chemical process.

This paper is organized as follows: Section II gives a brief review of KPCA technique and the fuzzy C-means clustering method. Section III defines the principles of the proposed FCM-KPCA approach for fault detection. The Tennessee Eastman Process is given in section IV to demonstrate the effectiveness of the proposed approach. Section V concludes the topics discussed in this paper and proposes future research work.

II. Preliminaries

A. KPCA

PCA is a simple linear transformation approach that compresses high-dimensional data with minimum loss of data information. In general, PCA can only be effectively performed on a set of observations that vary linearly. When the variations are nonlinear, some nonlinear approaches are required. KPCA has been successfully developed for tackling the nonlinear problem in recent years. KPCA maps the input data x into a feature space F via a nonlinear mapping. According to Cover's theorem [23], the nonlinear data structure in the input space is more likely to be linear after high-dimensional nonlinear mapping, and so, a linear PCA can be performed in F.

Given a set of normalized training data $\{x_i\}_{i=1}^{n} \in \Re^m$ and $m << n$. Let Φ is a nonlinear mapping: $x \in \Re^m \rightarrow \Phi(x) \in F^a$, the training data are then extended into hyper-dimensional feature space F, where, the dimension of F, a, can be arbitrary large or even infinite. The sample covariance matrix in the in

Lamiaa M. Elshenawy is with the Industrial Electronics and Control Engineering Department, Faculty of Electronic Engineering, Menoufiya University, 32952 Menouf, Egypt (email:lamia.elshennawy@el-eng.menofia.edu.eg).

Tarek A. M. Mohamed is with the Industrial Electronics and Control Engineering Department, Faculty of Electronic Engineering, Menoufiya University, 32952 Menouf, Egypt (email: tarek.mohamed@el-eng.menofia.edu.eg).

the feature space can be computed as:

$$S_\Phi = \frac{1}{n}\sum_{i=1}^{n}\overline{\Phi}(x_i)\overline{\Phi}(x_i)^T \quad (1)$$

where $\overline{\Phi}(x_i) = \Phi(x_i) - \mathrm{m}_\Phi$, and $m_\Phi = 1/n\sum_{i=1}^{n}\Phi(x_i)$ is the sample mean in the feature space. The eigenvalues $\lambda \geq 0$ and the associated eigenvectors $v \neq 0 \in F$ are then computed by solving the following eigenvalue problem in the feature space:

$$\lambda v = S_\Phi v = \left(\frac{1}{n}\sum_{i=1}^{n}\overline{\Phi}(x_i)\overline{\Phi}(x_i)^T\right) v \quad (2)$$

By the same argument in the input space, the solutions υ lie in the span of $\{\overline{\Phi}(x_i)\}_{i=1}^{n}$. Thus, we can consider the equivalent equations:

$$S_\Phi v = \frac{1}{n}\sum_{i=1}^{n}(\overline{\Phi}(x_i)\cdot v)\overline{\Phi}(x_i) \quad (3)$$

$$\lambda(\overline{\Phi}(x_i)\cdot v) = (\overline{\Phi}(x_i)\cdot S_\Phi v) \quad (4)$$

and the coefficients $\{\beta_i\}_{i=1}^{n}$ such that:

$$v = \sum_{i=1}^{n}\beta_i\overline{\Phi}(x_i) \quad (5)$$

To calculate the dot products $(\overline{\Phi}(x_i)\cdot\overline{\Phi}(x_j))$ in the feature space, kernel function $k(x_i, x_j)$ can be used [23]. Some of the most widely used kernel functions are: Gaussian radial basis kernel: $k(x, y) = \exp(-\|x-y\|^2/\sigma)$; polynomial kernel $k(x, y) = (x.y)^d$; inhomogeneous polynomial kernel: $k(x, y) = ((x.y) + \omega)^d$; sigmoid kernel: $k(x, y) = tanh(b_0 (x.y) + b_1)$, where $\|.\|$ is l_2-norm; σ, d, ω, b_0, and b_1 have to be specified [24]. The specific choice of kernel function implicitly defines the form of mapping and the feature space. By defining a kernel matrix $\overline{K}_{ij} = (\overline{\Phi}(x_i).\overline{\Phi}(x_j)) \in \Re^{n\times n}$, the eigenvalue problem can be represented by the following simple form:

$$\lambda\beta = (1/n)\overline{K}\beta \quad (6)$$

Where $\beta = [\beta_1,...,\beta_n]$

The centered kernel matrix $\overline{K}_{ij}$ is easily calculated using $K_{ij} = (\overline{\Phi}(x_i).\overline{\Phi}(x_j)) : \overline{K} = K - 1_n K - K1_n + 1_n K1_n$, where $1_n = (1/n)_{n\times n}$. To ensure the normality of the eigenvector in the feature space, $\|v\|^2 = 1$ the calculated eigenvector β should be scaled by $\|\beta\|^2 = 1/(n\lambda_k)$.

The KPCA-based fault detection is similar to linear PCA. The monitoring indices, i.e. Hotelling's T^2 and Q in the feature space can be used for process monitoring. There are numerous methods for determining the number of significant PCs l, including cumulative percent variance, the scree test, average eigenvalues, and the variance of reconstruction error [25]. In general, the number of l selected for KPCA is larger than that of linear PCA because KPCA extracts the major PCs from the infinite high-dimensional feature space whereas linear PCA extracts the major PCs from the relatively small input space. The KPCA has the potential to utilize more PCs to code structure rather than noise. Hence, the KPCA outperforms linear PCA in denoising if a sufficient large number of PCs is selected [5].

After constructing the PCA model in the feature space F, the KPCA score vector $t \in \Re^l$ for a new sample $x \in \Re^m$; $t = [t_1,..., t_l]$ Can be given by:

$$t_k = (v_k.\overline{\Phi}(x)) = \sum_{i=1}^{n}\beta_i^k(\overline{\Phi}(x_i).\overline{\Phi}(x)) = \sum_{i=1}^{n}\beta_i^k\overline{K}(x_i, x) \quad (7)$$

for k=1, … ,l.

In the monitoring models, the changes in the score variables reflect the changes in the underlying process behavior. The variation within the KPCA model is measured by Hotelling's T^2 that is the sum of the normalized squared scores as follows:

$$T^2 = t^T\Lambda_l^{-1}t \quad (8)$$

Where $\Lambda_l \in \Re^{l\times l}$ denotes the diagonal matrix of significant eigenvalues. This statistic follows F distribution [26] and the confidence limit, T_δ^2 is given by:

$$T_\delta^2 = \frac{l(n-1)}{n-l}F_{l,n-1,\delta} \quad (9)$$

Where δ is the confidence level.

The variability that breaks the normal process correlation is measured by Q statistic [27]:

$$Q = \overline{K}(x, x) - t^T t \quad (10)$$

Where $\overline{K}(x, x) \in \Re^{1\times 1}$ is the mean centered test kernel vector.

$$\overline{K}(x, x) = K(x, x) - 21_t^T K(X, x) + 1_t^T K1_t \quad (11)$$

with $1_t = (1/n)[1...1]^T \in \Re^n$. $K(X, x)$ represents the kernel vector constructed $\{x_i\}_{i=1}^{n}$ and x

$$K(X, x) = [K(x_1, x), K(x_2, x),..., K(x_n, x)]^T \quad (12)$$

The confidence limit of Q is:

$$Q_\delta = \vartheta_1\left[\frac{c_\delta\sqrt{2\vartheta_2 h_0^2}}{\vartheta_1} + \frac{\vartheta_2 h_0(h_0-1)}{\vartheta_1^2} + 1\right]^{1/h_0} \quad (13)$$

where $\vartheta_i = \sum_{i=l+1}^{n}\lambda_i^j$, j=1,2,3 and $h_0 = 1 - (2\vartheta_1\vartheta_3)/(3\vartheta_2^2)$

B. Fuzzy C-means clustering method

The idea behind FCM clustering method is to select the center of data representing the characteristics in every cluster.

Suppose the training data $\{x_i\}_{i=1}^{n} \in \Re^m$ and the problem is to determine a set of c-centroids $\{L_h\}_{h=1}^{c}$ so as to minimize the sum-of-squares criterion:

$$Minimize \quad J_c = \sum_{i=1}^{n}\sum_{h=1}^{c} (\mu_{ih})^p d^2(x_i, L_h) \tag{14}$$

where p > 1 is a weighting exponent that determines the fuzziness degree. $\mu_{ih} \subset \Omega$ represents the membership degree of the data point x_i to a cluster A_h, where $A = \{A_1, A_2, \ldots, A_c\}$ is the fuzzy partition. $d(x_i, L_h)$ is the distance between the data point x_i and cluster A_h. This distance is usually calculated by the Euclidean distance as:

$$d(x_i, L_h) = \left\| x_i - L_h \right\|^2 \tag{15}$$

The optimal fuzzy partition can be calculated according to the following algorithm, where the objective function is successively minimized with respect to Ω and **L**.

1. Let a data set x_i and $p > 1$. Initialize the partition matrix $\Omega(0)$ and set the iterative index r =0
2. Calculate at $(r = 0)$ or update at $(r > 0)$ the centroids $L_h(r)$, which minimize the objective function J_c

$$L_h(r) = \frac{\sum_{i=1}^{n} \mu_{ih}^p x_i}{\sum_{i=1}^{n} \mu_{ih}^p} \tag{16}$$

3. Update $\Omega(r)$ by:

$$\mu_{ih} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d^2(x_i, L_h)}{d^2(x_i, L_j)} \right)^{1/(p-1)}} \tag{17}$$

$h, j = 1,2,\ldots,c$ and $i = 1,\ldots,n$

4. Calculate the distance $\Delta = \left\| \Omega(r+1) - \Omega(r) \right\|$. If

$$\begin{cases} \Delta < \varepsilon & stop \\ \Delta \geq \varepsilon & go \quad to\ step\ 2 \end{cases}$$

where $\varepsilon > 0$.

III. FCM-KPCA Based Fault Detection

The concept of the proposed FCM-KPCA approach in this section is to use the FCM method to search about a subset of the samples in the input space whose mappings in the feature space are sufficient to express all of the data. As well known, the dimension of KPCA model depends on the number of training data as aforementioned in section II-A, which increases the computational cost in the training phase that is used to perform the eigen-decomposition. This computation cost will extend to the on-line phase to calculate the monitoring indices that can be possibly very high.

According to the mathematical analysis in section II of both KPCA and FCM clustering method, the proposed monitoring approach is detailed as follows: The sample covariance matrix in the feature space is modified into:

$$S_{\Phi_c} = \frac{1}{c}\sum_{i=1}^{c} \overline{\Phi}(L_i)\overline{\Phi}(L_i)^T \tag{18}$$

Therefore, the eigenvectors v_c lie in the span of $\{\overline{\Phi}(L_i)\}_{i=1}^{c}$, can be calculated according to:

$$S_{\Phi_c} v_c = \frac{1}{c}\sum_{i=1}^{c} (\overline{\Phi}(L_i) \cdot v_c)\overline{\Phi}(L_i) \tag{19}$$

$$v_c = \sum_{i=1}^{c} \beta_{c_i} \overline{\Phi}(L_i) \tag{20}$$

The new centered kernel matrix can be rewritten as: $\overline{K}_{cij} = (\overline{\Phi}(L_i).\overline{\Phi}(L_j)) \in \Re^{c \times c}$. The score vector $t_c \in \Re^{lc}$ can then be calculated by:

$$t_{ck} = (v_{ck}.\overline{\Phi}(x)) = \sum_{i=1}^{c} \beta_{ci}^k (\overline{\Phi}(L_i).\overline{\Phi}(x)) = \sum_{i=1}^{c} \beta_{ci}^k \overline{K}_c(L_i, x) \tag{21}$$

where l_c is the number of significant PCs in the feature space of FCM-KPCA, $k = 1, \cdots, l_c$. The monitoring indices will be:

$$T_c^2 = t_c^T \Lambda_{lc}^{-1} t_c = [t_{c1}, \ldots, t_{c_{lc}}] \Lambda_{l_c}^{-1} [t_{c_1}, \ldots, t_{c_{lc}}]^T \tag{22}$$

$$Q_c = \overline{K}_c(x, x) - t_c^T t_c \tag{23}$$

$$\overline{K}_c(x, x) = K_c(x, x) - 2 1_{tc}^T K_c(L, x) + 1_{tc}^T K_c 1_{tc} \tag{24}$$

with $1_{tc} = (1/c)[1 \ldots 1]^T \in \Re^c$. $K_c(L, x)$ represents the kernel vector constructed $\{L_h\}_{h=1}^{c}$ and x.

$K_c(L, x) = [K_c(L_1, x), K_c(L_2, x), \ldots, K_c(L_c, x)]^T$. The control limits of the monitoring indices T_c^2, Q_c are as follows:

$$T_{c\delta}^2 = \frac{l_c(c-1)}{c - l_c} F_{lc,\, c-lc,\, \delta}$$

$$Q_{c\delta} = \vartheta_1 \left[\frac{c_\delta \sqrt{2\vartheta_2 h_0^2}}{\vartheta_1} + \frac{\vartheta_2 h_0 (h_0 - 1)}{\vartheta_1^2} + 1 \right]^{1/h_0} \tag{25}$$

$\vartheta_i = \sum_{i=lc+1}^{c} \lambda_i^j$, $j = 1,2,3$ and $h_0 = 1 - (2\vartheta_1\vartheta_3)/(3\vartheta_2^2)$

According to the above analysis, the proposed approach and the original KPCA model can be compared in terms of the computational cost. This comparison is based on the number of floating point operations (flops). It should be noted that the most computational load is consumed in calculating the eigenpairs of the covariance matrix in the feature space which consumes $(22n^3 + 4n^2 + 12n + 1)$ or $O(n^3)$ based on singular value decomposition technique (SVD) [28]. Whereas, this cost can be reduced to $(22c^3 + 4c^2 + 12c + 1)$ or $O(c^3)$, $c < n$ in the proposed FCM-KPCA approach. Moreover, the computation cost of calculating the score vector t that used in determining the monitoring statistics, i.e. T^2 and Q depends on the number of the clusters c in FCM-KPCA instead of the training data n in case of KPCA.

The outline of the proposed FCM-PCA approach can be summarized as in the following: It includes two phases. First, constructing the normal operating condition model which is built off-line. Second, applying the monitoring model on-line.

(1) Constructing the normal operating condition (NOC) model:

1) Collect the historical data during normal operation and normalize them to zero mean and unit variance;
2) Calculate the normalized c-centroids $\{L_h\}_{h=1}^{c}$ using FCM method
3) Compute the centered kernel matrix K $\in \Re^{c\times c}$ and its eigenpairs;
4) Calculate the control limits of the monitoring statistics, i.e. Hotelling's T_c^2 and Q_c as in (25).

(2) On-line monitoring model:

1) Normalize a new test data vector *x* to zero mean and unit variance;
2) Determine the score vector tc for *x* as in (21);
3) Calculate the monitoring indices (T_c^2 or Q_c);
4) Check whether T_c^2 and/or Q_c exceeds the corresponding control limits $T_{c\delta}^2$ and $Q_{c\delta}$, respectively. If so, the process is out-of statistical-control due to faults and the alarm is released, otherwise the process is in-statistical-control.
5) Go to step 1.

In this article, the PCA and the proposed FCM-KPCA approach are compared in terms of the sensitivity for all possible faults of a process. This can be quantified by calculating the missed detection rates (MDR) for the processes faults of the testing data sets according to the following equation:

$$MDR = \frac{\text{undetected samples}}{\text{faulty samples}}\% \quad (26)$$

This criterion is used in the next section to measure the affectivity of the proposed approach, i.e. FCM-KPCA over PCA.

IV. Simulation Results

In this section, the PCA and the proposed FCM-KPCA monitoring methods are applied to the well-known Tennessee Eastman (TE) chemical process. TE process was first introduced by [29] and has been widely used for testing various process monitoring methods [30], [31], and [32]. The process is constructed by five major operation units: a reactor, a product condenser, a vapor-liquid separator, a recycle compressor and a product stripper as shown in Fig. 1. Two products are created from four reactants by means of two irreversible and exothermic reactions. It includes 52 input variables, which are composed of 41 measured and 11 manipulated variables, respectively. The details on the process description are well explained in [33]. The simulation data can be downloaded from the website http://brahms.scs.uiuc.edu. The total number of observations in the training dataset is 500. A set of 21 programmed faults is listed in Table I. The testing data sets for each fault are composed of 960 observations and all faults were introduced from sample 160.

TABLE I
Process Faults In Tennessee Eastman Process

Case	Process variable Type	Type
1	A/C feed ratio, B composition constant	step
2	B composition, A/C ration constant	step
3	D feed temperature	step
4	Reactor cooling water inlet temperature	step
5	Condenser cooling water inlet temperature	step
6	A feed loss	step
7	C header pressure loss-reduced availability	step
8	A,B,C feed composition	Random variation
9	D feed temperature	Random variation
10	C feed temperature	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics Reaction kinetics	Random variation
14	Reactor cooling water valve Sticking	Sticking
15	Condenser cooling water valve Sticking	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	The valve for stream 4 was fixed at the steady state position	Constant position

A. Fault detection results

In our study, a Gaussian radial basis kernel function with a

width σ empirically chosen is considered to construct the FCM-KPCA model. The number of clusters c was chosen as .

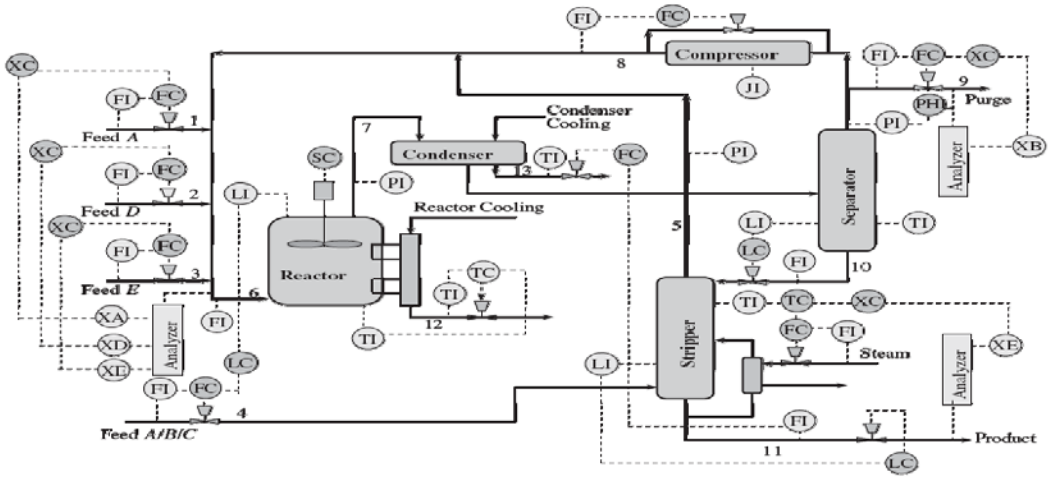

Fig. 1 Tennessee Eastman challenge process schematic diagram

100. The numbers of significant PCs were selected using CPV method as 40, 39 in PCA and FCM-KPCA, respectively. The number of clusters c was chosen as 100. The numbers of significant PCs were selected using CPV method as 40, 39 in PCA and FCM-KPCA, respectively. To illustrate the strengths and weaknesses of each statistic in both monitoring methods, faults 7 and 14 are selected. The monitoring results for these faults cases are illustratively shown in Figs. 2-5 using FCM-KPCA approach in comparison with PCA.

In the case of fault 7, the C header pressure loss-reduced availability is changed step by step. The monitoring results based on PCA and FCM-KPCA are shown in Fig. 2 and 3, respectively. PCA can detect the faults from sample 160 but the Q statistic fails to detect this fault at about samples 400, which raises the missed detection rate to 34.6683%. On the other hand, the proposed FCM-KPCA successes in detecting this fault at sample 160 and continued until the end of the simulation without missed detection samples.

In the case of fault 14, the reactor cooling water valve is sticked. Both PCA and FCM-KPCA could detect the fault from about sample 160 (Figs. 4-5). However, the Q-statistic of PCA could not detect the fault properly. In contrast to PCA, both statistics of FCMKPC, i.e. Tc^2 and Q_c showed high performance in detecting this fault. To detail the performance of PCA and FCM-KPCA, the corresponding missed detection rates for all faults 1-21 are computed and listed in Table II. Obviously, for most cases, the fault detection rates based on FCM-KPCA are higher than PCA. It should be noted that the faults 3, 9, and 15 have no observable change in the mean or the variance. In other words, these faults are unobservable from the testing data set which yields to high missed detection rates for most monitoring statistics approaches as in [33]. However, the proposed FCMKPCA approach could decrease the average missed detection rates for these faults from 88.1936% in the case of PCA's statistics to about 40.9262%. The results prove the efficiency of FCM-KPCA in monitoring nonlinear processes. In addition to that, it could decrease the computational cost of the original KPCA model.

V. Conclusion

This article has studied the application of KPCA for monitoring nonlinear processes. The contribution of this article is the development of a numerically efficient KPCA monitoring approach. FCM clustering technique is used to classify the training data into c-clusters. The centroids of these clusters are mapped to an infinite-dimensional feature space to build FCM-KPCA approach. To evaluate the proposed FCM-KPCA approach, the paper presented the analysis of recorded data from a Tennessee Eastman chemical process. The evaluation and comparison of the monitoring statistics are based on criteria that quantify the process monitoring performance. The simulation results confirmed that the proposed monitoring approach with a reduced computational cost can still effectively capture the nonlinear relationships in process variables. Therefore, it has been effective in detecting abnormal events with less missed detection rates than PCA.

Future work will concentrate on how to improve the ability of the proposed approach for fault detection in nonlinear time-varying industrial processes.

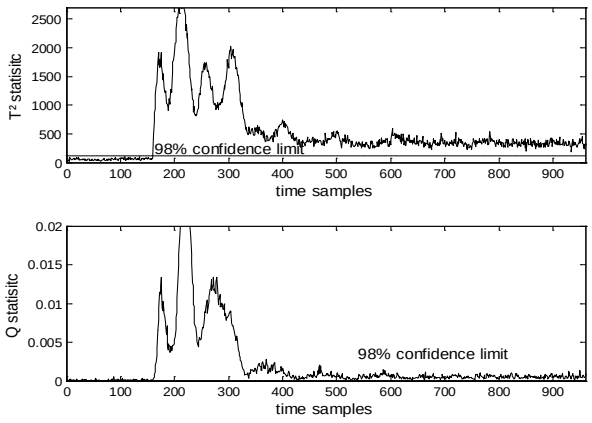

Fig. 2 PCA fault detection results for Tennessee Eastman process (Fault 7)

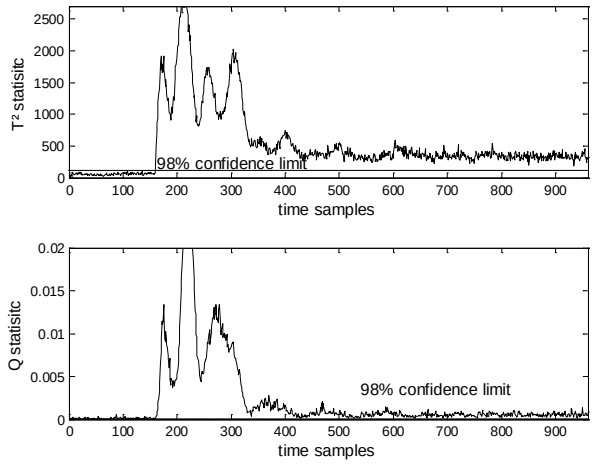

Fig. 3 KPCA-FCM fault detection results for Tennessee Eastman process (Fault 7)

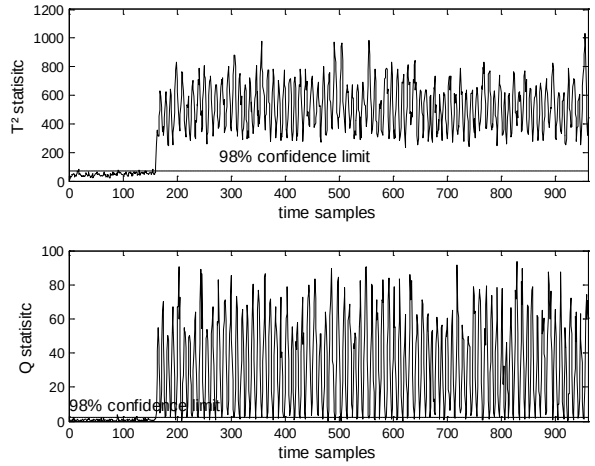

Fig. 4 PCA fault detection results for Tennessee Eastman process (Fault 14)

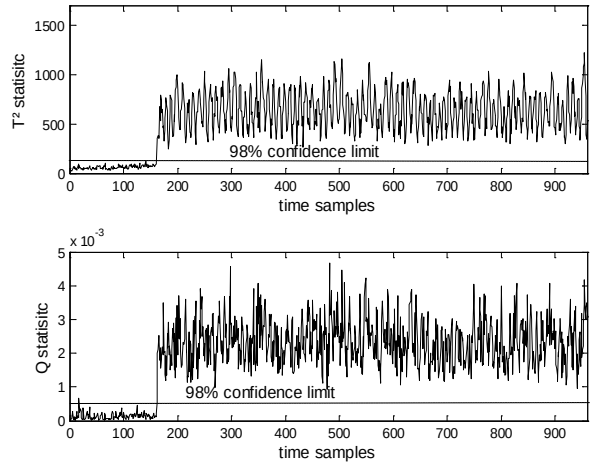

Fig. 5 KPCA-FCM fault detection results for Tennessee Eastman process (Fault 14)

TABLE II
MISSES DETECTION RATES OF PCA AND FCM-KPCA IN THE TENNESSEE EASTMAN PROCESS

Cases	PCA		FCM-KPCA	
	T^2	Q	T^2	Q
1	0.1252	0.1252	0.2503	0
2	1.2516	1.6270	0.6258	0.1252
3	85.3567	93.1164	46.0576	34.1677
4	0.1252	0.7509	0	2.1277
5	65.8323	21.5269	36.2954	0.5006
6	0.5006	0	0.1252	0
7	0	34.6683	0	0
8	1.7522	7.1339	1.1264	0.6258
9	88.1101	90.8636	47.5594	33.9174
10	40.0501	28.7860	20.0250	6.5081
11	20.9011	33.4168	7.5094	10.3880
12	0.7509	6.6333	0	0.2503
13	4.2553	4.1302	2.6283	1.8773
14	0	8.3855	0	0
15	80.7259	90.9887	48.9362	34.9186
16	53.6921	28.2854	20.2753	4.3805
17	5.5069	3.6295	3.7547	1.6270
18	8.1352	9.3867	3.6295	2.7534
19	75.3442	30.1627	32.9161	10.0125
20	38.1727	29.5369	13.5169	12.0150
21	45.5569	39.2991	25.9074	22.4030

REFERENCES

[1] D. Xiaogang and T. Xuemin, "Nonlinear process monitoring using dynamic sparse kernel classifier," *Procedia Engineering*, vol. 29, pp. 295–300, 2012. http://dx.doi.org/10.1016/j.proeng.2011.12.710

[2] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, 1995.

[3] N. Cristianini and J. Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, England: Cambridge University Press, 2000. http://dx.doi.org/10.1017/CBO9780511801389

[4] S. Wang, J. Yu, E. Lapira, and J. Lee, "A modified support vector data description based novelty detection approach for machinery components," *Applied Soft Computing*, vol. 13, no. 2, pp. 1193–1205, 2013. http://dx.doi.org/10.1016/j.asoc.2012.11.005

[5] J. M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel component analysis," *Chemical Engineering Science*, vol. 59, pp. 223–234, 2004. http://dx.doi.org/10.1016/j.ces.2003.09.012

[6] Z. G. Zhao and F. Liu, "On-line nonlinear process monitoring using kernel principal component analysis and neural network," *Springer-Verlag Berlin Heidelberg*, pp. 945–950, 2006.

[7] Y. Zhang and S. Qin, "Improved nonlinear fault detection technique and statistical analysis," *AIChE Journal*, vol. 54, no. 12, pp. 3207–3220, December 2008. http://dx.doi.org/10.1002/aic.11617

[8] Y. W. Zhang, H. Zhou, and S. J. Qin, "Decentralized fault diagnosis of large-scale processes using multiblock kernel principal component analysis," *Acta Automatica Sinica*, vol. 36, no. 4, pp. 593–579, 2010. http://dx.doi.org/10.3724/SP.J.1004.2010.00593

[9] X. Deng, X. Tian, and S. Chen, "Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis," *Chemometrics and Intelligent Laboratory Systems*, vol. 127, pp. 195–209, 2013. http://dx.doi.org/10.1016/j.chemolab.2013.07.001

[10] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *J. Mach. Learning Res.*, vol. 2, pp. 97–123, 2001.

[11] Y. Zhang and C. Ma, "Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS," *Chemical Engineering Science*, vol. 66, pp. 64–72, 2011. http://dx.doi.org/10.1016/j.ces.2010.10.008

[12] Y. Zhang, L. Zhang, and R. Lu, "Fault identification of nonlinear processes," *Ind. Eng. Chem. Res.*, vol. 52, no. 34, pp. 12 072–12 081, 2013.

[13] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385–2404, 2000. http://dx.doi.org/10.1162/089976600300014980

[14] J. Yang, A. Frangi, J.-Y.Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005. http://dx.doi.org/10.1109/TPAMI.2005.33

[15] X. Zhang, X. Zhao, and H. Shao, "Nonlinear real-time process monitoring and fault diagnosis based on principal component analysis and kernel fisher discriminant analysis," *Chem. Eng. Technol.*, vol. 30, no. 9, pp. 1203–1211, 2007. http://dx.doi.org/10.1002/ceat.200600410

[16] Z. Zhu and Z. Song, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Systems with Applications*, vol. 38, pp. 6895–6905, 2011. http://dx.doi.org/10.1016/j.eswa.2010.12.034

[17] Y. Zhang, "Fault detection and diagnosis of nonlinear processes using improved kernel independent component analysis (KICA) and support vector machine (SVM)," *Ind. Eng. Chem. Res.*, vol. 47, pp. 6961–6971, 2008. http://dx.doi.org/10.1021/ie071496x

[18] Q. Jiang and X. Yan, "Non-gaussian chemical process monitoring with adaptively weighted independent component analysis and its applications,"*Journal of Process Control*, vol. 23, no. 9, pp. 1320–1331, 2013. http://dx.doi.org/10.1016/j.jprocont.2013.09.008

[19] J. Fan and Y. Wang, "Fault detection and diagnosis of non-linear non-gaussian dynamic processes using kernel dynamic independent component analysis," *Information Sciences, In Press, Corrected Proof*, 2013.

[20] C. Zhao, F. Wang, and Y. Zhang, "Nonlinear process monitoring based on kernel dissimilarity analysis," *Control Engineering Practice*, vol. 17, pp. 221–230, 2009. http://dx.doi.org/10.1016/j.conengprac.2008.07.001

[21] P. Cui, J. Li, and G. Wang, "Improved kernel principal component analysis for fault detection," *Expert Systems with Applications*, vol. 34, pp. 1210–1219, 2008. http://dx.doi.org/10.1016/j.eswa.2006.12.010

[22] X. Rui and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005. http://dx.doi.org/10.1109/TNN.2005.845141

[23] B. Schölkopf, A. J. Smola, and K. M¨uller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, p. 1299, 1998. http://dx.doi.org/10.1162/089976698300017467

[24] Y. Zhang, "Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM," *Chemical Engineering Science*, vol. 64, pp. 801–811, 2009. http://dx.doi.org/10.1016/j.ces.2008.10.012

[25] X. Liu, U. Kruger, T. Littler, L. Xie, and S. Wang, "Moving window kernel pca for adaptive monitoring of nonlinear processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 96, no. 2, pp. 132–143, 2009. http://dx.doi.org/10.1016/j.chemolab.2009.01.002

[26] J. E. Jackson, *A User's Guide to Principal Components*. New York: John Wiley & Sons, Inc, 1991. http://dx.doi.org/10.1002/0471725331

[27] S. J. Qin, "Statistical process monitoring: basics and beyond," *J. Chemometrics*, vol. 17, pp. 480–502, 2003. http://dx.doi.org/10.1002/cem.800

[28] G. H. Golub and C. V. Loan, *Matrix computations, 3rd edition*. London: Johns Hopkins Press Ltd., 1996.

[29] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245– 255, 1993. http://dx.doi.org/10.1016/0098-1354(93)80018-I

[30] L. Biegler, T. Jockenhvel, and A. Whter, "Dynamic optimization of the tennessee eastman process using the optcontrolcentre," *Computers and Chemical Engineering*, vol. 18, pp. 383–413, 2003.

[31] A. Singhal and D. Seborg, "Evaluation of a pattern matching method for the tennessee eastman challenge process," *Journal of Process Control*, vol. 16, pp. 601–613, 2006. http://dx.doi.org/10.1016/j.jprocont.2005.10.005

[32] Z. Ge and Z. Song, "Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors," *Ind. Eng. Chem. Res.*, vol. 46, pp. 2054–2063, 2007. http://dx.doi.org/10.1021/ie061083g

[33] L. H. Chiang, E. Russel, and R. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London: Springer, 2001. http://dx.doi.org/10.1007/978-1-4471-0347-9