

# Monitoring Nonlinear and Non-Gaussian Processes Using Gaussian Mixture Model Based Weighted Kernel Independent Component Analysis

Lianfang Cai, Xuemin Tian, and Sheng Chen, *Fellow, IEEE*

**Abstract**—Kernel independent component analysis (KICA) is widely regarded as an effective approach for nonlinear and non-Gaussian process monitoring. However, the KICA-based monitoring methods treat every kernel independent component (KIC) equally and cannot highlight the useful KICs associated with fault information. Consequently, fault information may not be explored effectively which may result in degraded fault detection performance. To overcome this problem, we propose a new nonlinear and non-Gaussian process monitoring method using Gaussian mixture model (GMM) based weighted kernel independent component analysis (WKICA). Specifically, in WKICA, GMM is firstly adopted to estimate the probabilities of the KICs extracted by KICA. The significant KICs embodying the dominant process variation are then discriminated based on the estimated probabilities and assigned with larger weights to capture the significant information during on-line fault detection. A nonlinear contribution plots method is also developed based on the idea of sensitivity analysis to help identifying the fault variables after a fault is detected. Simulation studies conducted on a simple four-variable nonlinear system and the Tennessee Eastman benchmark process demonstrate the superiority of the proposed method over the conventional KICA-based method.

**Index Terms**—Process monitoring, fault detection, kernel independent component analysis, Gaussian mixture model, fault identification, contribution plots

## I. INTRODUCTION

Modern industrial processes are large scale and highly complex. Efficient and reliable process monitoring plays a key role in ensuring process safety and product quality. With wide applications of distributed control systems and measurement technology, large amounts of data are collected in today's process industry, which facilitate rapid development of data-driven multivariate statistical methods [1]–[7] for process monitoring. Principal component analysis (PCA), as a classical multivariate statistical method, has gained much attention from both academia and industry [8], [9]. It projects the original high-dimensional process variables onto a low-dimensional space that retains most of the original variance to obtain a smaller set of the uncorrelated latent variables called principal components (PCs). Many extensions [10]–[12] of PCA have been developed to improve process monitoring performance by taking different process characteristics into consideration. However, PCA only considers the second-order statistic and

cannot make use of the higher-order statistical information in non-Gaussian process data [13], [14]. Since process data are usually non-Gaussian distributed as a result of nonlinearity, operating condition shifts or other reasons [15], this limitation of PCA may result in inadequate feature extraction in non-Gaussian processes. Moreover, in PCA-based fault detection, the assumption that the process data obey a multivariate Gaussian distribution is required for determining the confidence limits of the Hotelling's T-squared ( $T^2$ ) and corresponding squared prediction error (SPE) monitoring statistics. Since the confidence limits derived based on the Gaussian assumption are inappropriate for monitoring non-Gaussian processes, the resulting fault indications from the constructed  $T^2$  and SPE monitoring charts may be misleading.

More recently, independent component analysis (ICA), as a newly emerging multivariate statistical approach, has exhibited tremendous potential for non-Gaussian process monitoring [4], [5], [16], [17]. In comparison to PCA, ICA can effectively utilize the higher-order information in non-Gaussian process data and extracts mutually independent latent variables known as independent components (ICs) from the original process variables. Thus, ICs can reveal more useful information than PCs from non-Gaussian data [13], which is usually the case for industrial data. In this sense, ICA can be regarded as a useful extension of PCA. In [18], Kano *et al.* applied ICA to extract the ICs from process data and monitored each IC for detecting abnormal operating conditions. Their application results show the superior monitoring performance of ICA over PCA. To account for process dynamic characteristics, Odiwei and Cao [19] proposed a state-space ICA method, which adopts the canonical variate analysis to construct a state space by performing the dynamic whitening and then applies ICA in the obtained state space to extract the ICs. For monitoring complex processes with inherent system uncertainty and multiple operating conditions, Rashid and Yu [20] developed an adaptive ICA method based on the hidden Markov model.

Although ICA has demonstrated its effectiveness in non-Gaussian process monitoring, it is a linear statistical method, requiring the assumption that process data have linear structure. However, in industrial environments, the collected process data are usually nonlinear. Therefore, ICA may fail to conduct effective and adequate feature extraction from nonlinear process data, which may lead to unsatisfactory monitoring performance. Lee *et al.* [14] firstly introduced the kernel independent component analysis (KICA) of [21] to tackle the nonlinear process monitoring problem, and demonstrated that the fault detection performance achieved by KICA is better than that of ICA. Essentially, KICA is an integration of kernel

L. Cai and X. Tian are with College of Information and Control Engineering, China University of Petroleum, Qingdao 266580, China (E-mails: cailianfang@163.com, tianxm@upc.edu.cn).

S. Chen is with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: sqc@ecs.soton.ac.uk), and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia

This work was supported by the National Natural Science Foundation of China (Grant nos. 61273160 and 61403418).

PCA (KPCA) with ICA. Basically, it first projects the original nonlinear process data onto a high-dimensional linear feature space, executes the whitening operation in this linear feature space using KPCA, and then employs ICA to extract the kernel ICs (KICs) in the KPCA-whitened space. Because of the ability in dealing with nonlinear problems frequently encountered in process monitoring, KICA has become a hot research topic in recent years. Taking both the process nonlinearity and multimodality into consideration, Zhang *et al.* [4] introduced the Kronecker product to modify the monitoring matrices, and proposed a multimode KICA method for process monitoring. Fan *et al.* [22] proposed a filtering KICA-PCA method to improve the process monitoring performance, which applies genetic algorithm to determine the kernel parameter and adopts the exponentially weighted moving average technique to filter the monitoring statistics. To solve the monitoring problem of nonlinear batch processes, Tian *et al.* [23] proposed a multiway KICA method, which unfolds the three-way dataset into the two-way one and then selects representative feature samples to construct the KICA monitoring model.

It can be seen that KICA has been utilized as an effective means for nonlinear and non-Gaussian process monitoring. But the current KICA-based monitoring methods seldom investigate the significance of different KICs to process monitoring. Since individual KICs have different degrees of importance, in terms of revealing the process information, how to appropriately make use of the extracted KICs according to their related importance is vital for efficient and reliable process monitoring. Unfortunately, the existing KICA-based monitoring methods simply use the extracted KICs with equal weights, and they fail to properly consider the different importance degrees of the different KICs. Consequently, these existing methods cannot give sufficient attention to the significant KICs associated with useful fault information, which will inevitably cause the undesirable effect that the fault information may be submerged in the insignificant KICs. As a result, the occurring fault may not be detected timely. A main motivation of our work is based on the fact that the fault detection performance can be improved by emphasizing the important KICs that effectively reflect the process state. Another critical issue worthy noting is the fault identification after a fault is detected. The existing KICA-based monitoring methods generally take the fault detection as the primary target, but they rarely pay sufficient attention to the fault identification. However, after fault detection, fault identification is the crucial step to identify the fault variables, which is vital for guiding the repair of the detected fault. Therefore, in order to ensure the process return to the normal operating state fast and efficient, it is necessary to develop an effective nonlinear fault identification method naturally associated with KICA-based process monitoring, which however remains to be a challenging open problem.

Against this background, we propose a new monitoring method for nonlinear and non-Gaussian processes using a novel Gaussian mixture model (GMM) based weighted KICA (WKICA). As usual, nonlinear process data are projected onto a high-dimensional linear feature space and the KICs are extracted in the feature space by KICA. However, we employ the GMM [24] to estimate the probability of each obtained

KIC for measuring the individual KIC's importance. This enables us to assign different weight values to the extracted KICs according to their measured importance for highlighting the important process information when on-line fault detection is implemented. A further contribution is to propose a new nonlinear contribution plots method for the challenging fault identification problem, which is developed based on the idea of sensitivity analysis [25]. Two case studies, involving a simple four-variable nonlinear system and the Tennessee Eastman benchmark process, are used to demonstrate the effectiveness of the proposed nonlinear monitoring method.

## II. CONVENTIONAL MONITORING METHOD USING KICA

KICA contains two essential steps: 1) Projecting the nonlinear process data onto a high-dimensional linear feature space to obtain the feature data and whitening the feature data in the feature space using KPCA; 2) Performing ICA on the whitened data in the KPCA-whitened space to extract the KICs features.

Specifically, denote  $\mathbf{x}_i \in \mathbb{R}^m$ ,  $1 \leq i \leq n$ , as the normal operating data with  $m$  process variables and  $n$  samples. The nonlinear mapping  $\phi(\cdot) : \mathbb{R}^m \rightarrow \mathcal{F}$  projects the nonlinear data  $\mathbf{x}_i$  for  $1 \leq i \leq n$  in the original variable space onto the high-dimensional linear feature space  $\mathcal{F}$  to obtain the high-dimensional feature data  $\phi(\mathbf{x}_i) \in \mathcal{F}$ ,  $1 \leq i \leq n$ . Further denote  $\Phi = [\phi(\mathbf{x}_1) \phi(\mathbf{x}_2) \cdots \phi(\mathbf{x}_n)] \in \mathcal{F} \times \mathbb{R}^n$  and perform mean centering on  $\Phi$  in the feature space to acquire the zero-mean feature data  $\bar{\Phi} = [\bar{\phi}(\mathbf{x}_1) \bar{\phi}(\mathbf{x}_2) \cdots \bar{\phi}(\mathbf{x}_n)] \in \mathcal{F} \times \mathbb{R}^n$ . Then, KPCA can be adopted to conduct the whitening of  $\bar{\Phi}$ .

The covariance matrix  $C \in \mathcal{F} \times \mathcal{F}$  of the data  $\bar{\Phi}$  may be estimated by  $C = \frac{1}{n} \bar{\Phi} \bar{\Phi}^T$ . A straightforward idea would be to find the eigenvectors of  $C$  for obtaining the PCs in the feature space. However, the functional form of  $\phi(\cdot)$  is unknown, which makes it infeasible by the eigen-decomposition of  $C$  directly. Hence, 'kernel trick' is introduced to solve this problem. Define a Gram kernel matrix  $K \in \mathbb{R}^{n \times n}$  as

$$K = \Phi^T \Phi. \quad (1)$$

The  $i$ th-row and  $j$ th-column element of  $K$  can be written as

$$[K]_{i,j} = \phi^T(\mathbf{x}_i) \phi(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j), \quad (2)$$

where  $k(\mathbf{x}_i, \mathbf{x}_j)$  is the kernel function. The choice of  $k(\cdot, \cdot)$  implicitly determines the nonlinear mapping  $\phi(\cdot)$  and the corresponding high-dimensional feature space  $\mathcal{F}$ . A widely used kernel function is the Gaussian kernel given by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{c}\right), \quad (3)$$

where  $c$  is the kernel width. By using kernel function  $k(\cdot, \cdot)$ , the inner product of two high-dimensional feature data in the feature space can be calculated in the input space, without having to perform the nonlinear mapping  $\phi(\cdot)$  explicitly.

For the zero-mean feature data  $\bar{\Phi}$ , the corresponding Gram kernel matrix  $\bar{K} \in \mathbb{R}^{n \times n}$  can be obtained based on  $K$  by

$$\bar{K} = \bar{\Phi}^T \bar{\Phi} = K - \mathbf{1}_n K - K \mathbf{1}_n + \mathbf{1}_n K \mathbf{1}_n, \quad (4)$$

where  $\mathbf{1}_n \in \mathbb{R}^{n \times n}$  denotes the matrix whose elements are all equal to  $\frac{1}{n}$ . Let  $\lambda_i \in \mathbb{R}$ ,  $1 \leq i \leq n$ , be the eigenvalues of  $\bar{K}$

satisfying the condition  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ , and  $\beta_i \in \mathbb{R}^n$  for  $1 \leq i \leq n$  be the corresponding eigenvectors of  $\bar{\mathbf{K}}$ . The estimate of the covariance matrix  $\mathbf{C}$  can be expressed as

$$\mathbf{C} = (\bar{\Phi} \mathbf{H} \Lambda^{-1/2}) \frac{\Lambda}{n} (\bar{\Phi} \mathbf{H} \Lambda^{-1/2})^T, \quad (5)$$

where  $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \in \mathbb{R}^{n \times n}$  denotes the diagonal matrix with  $\lambda_1, \lambda_2, \dots, \lambda_n$  as its diagonal elements and  $\mathbf{H} = [\beta_1 \beta_2 \dots \beta_n] \in \mathbb{R}^{n \times n}$ . From (5), it can be observed that the  $a$  largest eigenvalues of the matrix  $\mathbf{C}$  are  $\frac{\lambda_1}{n}, \frac{\lambda_2}{n}, \dots, \frac{\lambda_a}{n}$ , and the corresponding eigenvector matrix  $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_a] \in \mathcal{F} \times \mathbb{R}^a$  of  $\mathbf{C}$  is as follows

$$\mathbf{V} = \bar{\Phi} \mathbf{H}_a \Lambda_a^{-1/2}, \quad (6)$$

where  $\Lambda_a = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_a\}$  and  $\mathbf{H}_a = [\beta_1 \beta_2 \dots \beta_a] \in \mathbb{R}^{n \times a}$ . Then, the kernel whitening matrix  $\mathbf{Q}_F \in \mathbb{R}^a \times \mathcal{F}$  can be obtained in the feature space  $\mathcal{F}$  as

$$\mathbf{Q}_F = \text{diag}\left\{\frac{\lambda_1}{n}, \frac{\lambda_2}{n}, \dots, \frac{\lambda_a}{n}\right\}^{-1/2} \mathbf{V}^T = \sqrt{n} \Lambda_a^{-1} \mathbf{H}_a^T \bar{\Phi}^T. \quad (7)$$

With the kernel whitening matrix  $\mathbf{Q}_F$ , the kernel whitened data  $\mathbf{Z} \in \mathbb{R}^{a \times n}$  can be deduced in the feature space as follows:

$$\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2 \dots \mathbf{z}_n] = \mathbf{Q}_F \bar{\Phi} = \sqrt{n} \Lambda_a^{-1} \mathbf{H}_a^T \bar{\mathbf{K}}. \quad (8)$$

where the covariance matrix of the kernel whitened data  $\mathbf{Z}$  is the  $a \times a$  identity matrix  $\mathbf{I}_a$ .

After obtaining the kernel whitened data by KPCA in the feature space, ICA is implemented on the kernel whitened data to extract the KICs. The maximum non-Gaussianity criterion is taken as the objective function of ICA, and the corresponding optimization problem is formulated as follows [26]:

$$\begin{aligned} \max_{\mathbf{u}_i} J(\mathbf{u}_i) &= \max_{\mathbf{u}_i} \left( \mathbb{E}\{G(\mathbf{u}_i^T \mathbf{z})\} - \mathbb{E}\{G(v)\} \right)^2, \\ \text{s.t. } \mathbb{E}\{(\mathbf{u}_i^T \mathbf{z})^2\} &= 1, \mathbf{u}_i^T \mathbf{u}_i = 1, \text{ and} \\ \mathbf{u}_i^T \mathbf{u}_{i-1} &= \mathbf{u}_i^T \mathbf{u}_{i-2} = \dots = \mathbf{u}_i^T \mathbf{u}_1 = 0, \end{aligned} \quad (9)$$

where  $\mathbf{z} \in \mathbb{R}^a$  denotes the vector of  $a$  whitened variables,  $\mathbf{u}_i \in \mathbb{R}^a$  denotes the  $i$ -th maximum non-Gaussian direction related to  $\mathbf{z}$ ,  $v$  is a Gaussian variable with zero mean and unit variance, and  $\mathbb{E}\{\cdot\}$  denotes the expectation operator, while  $G(\cdot)$  is a non-quadratic function that can be chosen as  $G(\mathbf{u}_i^T \mathbf{z}) = -\exp(-(\mathbf{u}_i^T \mathbf{z})^2/2)$ . The optimization algorithm for solving the optimization problem (9), as detailed in [26], has a fast convergence rate and can be carried out conveniently. Denote  $\mathbf{U} = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_a]^T$ . It can easily be verified that the matrix  $\mathbf{U}$  is an orthogonal matrix. Based on  $\mathbf{U}$ , the samples  $\mathbf{S} \in \mathbb{R}^{a \times n}$  of the KICs can be extracted according to

$$\mathbf{S} = [\mathbf{s}_1 \mathbf{s}_2 \dots \mathbf{s}_n] = \mathbf{U} \mathbf{Z}. \quad (10)$$

Let the current process data collected online be  $\mathbf{x}_t \in \mathbb{R}^m$ . The current sample  $\mathbf{s}_t \in \mathbb{R}^a$  of the KICs can be extracted as

$$\mathbf{z}_t = \mathbf{Q}_F \bar{\Phi}(\mathbf{x}_t) = \sqrt{n} \Lambda_a^{-1} \mathbf{H}_a^T \bar{\Phi}^T \bar{\Phi}(\mathbf{x}_t) = \sqrt{n} \Lambda_a^{-1} \mathbf{H}_a^T \bar{\mathbf{k}}_b \quad (11)$$

$$\mathbf{s}_t = \mathbf{U} \mathbf{z}_t, \quad (12)$$

where  $\bar{\Phi}(\mathbf{x}_t) \in \mathcal{F}$  denote the current zero-mean feature data in the feature space, and  $\mathbf{z}_t \in \mathbb{R}^a$  denote the current values of the whitened variables, while  $\bar{\mathbf{k}}_t = \mathbf{k}_t - \mathbf{K} \mathbf{1}_1 - \mathbf{1}_n \mathbf{k}_t + \mathbf{1}_n \mathbf{K} \mathbf{1}_1$

with the vector  $\mathbf{1}_1 \in \mathbb{R}^n$  whose elements are all equal to  $\frac{1}{n}$ , and  $\mathbf{k}_t = [k(\mathbf{x}_1, \mathbf{x}_t) \ k(\mathbf{x}_2, \mathbf{x}_t) \ \dots \ k(\mathbf{x}_n, \mathbf{x}_t)]^T$ .

To conduct fault detection based on the extracted KICs by KICA, two monitoring statistics are constructed [4], [14], [23]:

$$I_t^2 = (\mathbf{U}_d \mathbf{z}_t)^T \mathbf{U}_d \mathbf{z}_t = \mathbf{s}_{d,t}^T \mathbf{s}_{d,t}, \quad (13)$$

$$\begin{aligned} Q_t &= \mathbf{e}_t^T \mathbf{e}_t = (\mathbf{z}_t - \mathbf{U}_d^T \mathbf{U}_d \mathbf{z}_t)^T (\mathbf{z}_t - \mathbf{U}_d^T \mathbf{U}_d \mathbf{z}_t) \\ &= (\mathbf{U}_e \mathbf{z}_t)^T \mathbf{U}_e \mathbf{z}_t = \mathbf{s}_{e,t}^T \mathbf{s}_{e,t}, \end{aligned} \quad (14)$$

where  $d < a$  is the number of the dominant KICs,  $\mathbf{s}_{d,t} = [s_{1,t} \ s_{2,t} \ \dots \ s_{d,t}]^T = \mathbf{U}_d \mathbf{z}_t$  with  $\mathbf{s}_{d,t} \in \mathbb{R}^d$  contains the first  $d$  extracted KICs, called the dominant KICs at time  $t$ , and  $\mathbf{U}_d \in \mathbb{R}^{d \times a}$  consists of the first  $d$  rows of  $\mathbf{U}$ , while  $\mathbf{e}_t = \mathbf{z}_t - \mathbf{U}_d^T \mathbf{U}_d \mathbf{z}_t$  with  $\mathbf{e}_t \in \mathbb{R}^a$  is the residual vector of the KICA model,  $\mathbf{s}_{e,t} = [s_{d+1,t} \ s_{d+2,t} \ \dots \ s_{a,t}]^T = \mathbf{U}_e \mathbf{z}_t$  with  $\mathbf{s}_{e,t} \in \mathbb{R}^{a-d}$  contains the last  $a-d$  extracted KICs, called the excluded KICs at time  $t$ , and  $\mathbf{U}_e \in \mathbb{R}^{(a-d) \times a}$  consists of the last  $a-d$  rows of  $\mathbf{U}$ . The monitoring statistic  $I_t^2 \in \mathbb{R}$  is applied to detect the systematic part of the variation within the KICA model, while the monitoring statistic  $Q_t \in \mathbb{R}$  is used to detect the non-systematic part change in the residual of the KICA model [6], [13]. We determine the hyper dimension  $a$ , the number of dominant KICs  $d$ , and the kernel width  $c$  according to the empirical rule suggested in [14].

### III. PROPOSED NEW MONITORING METHOD

It is clear from (13) that in the existing KICA-based monitoring method, each dominant KIC plays equal role in constructing the monitoring statistic  $I_t^2$ . However, when a particular fault occurs, there usually exist some KICs in the set of dominant KICs which are specifically effective to detect this fault because they are sensitive to the occurring fault and can react to it fast, while the other dominant KICs may have slow reaction to the particular fault and are less beneficial for discovering it. Moreover, different faults have different relationships with the extracted KICs and the relationships between a specific fault and the KICs may also vary with time. If all the dominant KICs are adopted to detect faults with the same importance in the monitoring statistics at all time, the significant fault information in part of the dominant KICs may be suppressed or hidden by the information less relevant to the fault containing in the remaining KICs. This can lead to an unsatisfactory fault detection performance. Similarly, all the excluded KICs in the monitoring statistic  $Q_t$  of (14) are also treated to be equally important, which may reduce the effectiveness of this monitoring statistic for fault detection. Thus, when conducting on-line monitoring, an effective means for improving fault detection performance is to emphasize those KICs which contain useful fault information, while suppressing the KICs with insignificant information, according to the importance degrees of the KICs.

#### A. Fault detection based on WKICA

In order to make the useful fault information better reflected in the monitoring statistics, we assign larger weight values to the KICs which contain significant fault information but

smaller weight values to the KICs that contain no or insignificant fault information. With this weighting strategy, the two new monitoring statistics can be constructed as follows

$$WI_t^2 = (\mathbf{W}_{d,t} \mathbf{s}_{d,t})^T \mathbf{W}_{d,t} \mathbf{s}_{d,t} \quad \text{with } \mathbf{W}_{d,t} = \text{diag}\{w_{1,t}, w_{2,t}, \dots, w_{d,t}\}, \quad (15)$$

$$WQ_t = (\mathbf{W}_{e,t} \mathbf{s}_{e,t})^T \mathbf{W}_{e,t} \mathbf{s}_{e,t} \quad \text{with } \mathbf{W}_{e,t} = \text{diag}\{w_{d+1,t}, w_{d+2,t}, \dots, w_{a,t}\}, \quad (16)$$

where  $WI_t^2$  and  $WQ_t$  are respectively the improved versions of the monitoring statistics  $I_t^2$  and  $Q_t$  given in (13) and (14), while  $w_{i,t}$  is the weight that is determined according to the importance degree of the  $i$ th KIC at the time  $t$ .

To effectively measure the importance of the KICs at any time  $t$ , a probability evaluation method based on GMM is developed. Specifically, the two-Gaussian mixture [24] is adopted to fit the probability density function (PDF) of each extracted non-Gaussian KIC. GMM as a general and flexible density estimator is widely adopted for modeling non-Gaussian distributions from the datasets of realistic industrial processes [24], [27]–[30]. In particular, the GMM composed of mixture of two Gaussians, widely used for modeling PDFs in practice [24], [29], [30], offers a good trade-off between the complexity of estimating the parameters of the GMM and the capability of the GMM for accurately modeling underlying density. For the parameter estimation problem of this two Gaussian mixture, Santamaria *et al.* [24] specifically designed an improved expectation maximization (EM) algorithm.

Thus, for the  $i$ th extracted non-Gaussian KIC  $s_i$ , which has zero mean and unit variance, we use the following two-Gaussian mixture to fit the PDF of  $s_i$

$$p_i(s) = \xi_i g_0(s; \sigma_{i,1}^2) + (1 - \xi_i) g_0(s; \sigma_{i,2}^2), \quad 1 \leq i \leq a, \quad (17)$$

where  $g_0(\cdot; \sigma^2)$  denotes the zero-mean Gaussian PDF with variance  $\sigma^2$ , while the two variances  $\sigma_{i,1}^2$  and  $\sigma_{i,2}^2$  as well as the mixing proportion  $\xi_i$  satisfy the conditions:

$$\begin{cases} 0 < \xi_i < 1, \\ \xi_i \sigma_{i,1}^2 + (1 - \xi_i) \sigma_{i,2}^2 = \text{E}\{s_i^2\} = 1. \end{cases} \quad (18)$$

The parameters  $\xi_i$ ,  $\sigma_{i,1}^2$  and  $\sigma_{i,2}^2$  of this GMM are estimated using the improved EM algorithm of [24]. Denote the normal operating training samples of the KICs as  $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \dots \ \mathbf{s}_n]$  with  $\mathbf{s}_t = [s_{1,t} \ s_{2,t} \ \dots \ s_{a,t}]^T$  for  $1 \leq t \leq n$ . Then the parameters of the two Gaussian mixture are estimated iteratively according to:

$$\xi_i^{l+1} = \gamma \xi_i^l + \frac{1 - \gamma}{n} \sum_{t=1}^n \Theta_{i,t}^l \quad (19)$$

with

$$\Theta_{i,t}^l = \frac{\xi_i^l g_0(s_{i,t}; \sigma_{i,1}^2)^l}{\xi_i^l g_0(s_{i,t}; \sigma_{i,1}^2)^l + (1 - \xi_i^l) g_0(s_{i,t}; \sigma_{i,2}^2)^l}, \quad (20)$$

$$\sigma_{i,1}^{2(l+1)} = \gamma \sigma_{i,1}^{2l} + (1 - \gamma) \frac{\sum_{t=1}^n s_{i,t}^2 \Theta_{i,t}^l}{\sum_{t=1}^n \Theta_{i,t}^l}, \quad (21)$$

$$\sigma_{i,2}^{2(l+1)} = \frac{1 - \xi_i^{l+1} \sigma_{i,1}^{2(l+1)}}{1 - \xi_i^{l+1}}, \quad (22)$$

where  $(\cdot)^{l+1}$  denotes the value after the  $(l+1)$ th iteration, while  $0.8 < \gamma < 0.95$  is a smoothing factor. In our current study, we choose  $\gamma = 0.85$ . The iterative procedure is terminated when both  $|\xi_i^{l+1} - \xi_i^l|$  and  $|\sigma_{i,1}^{2(l+1)} - \sigma_{i,1}^{2l}|$  are smaller than a pre-defined threshold, e.g.,  $10^{-6}$ .

After the PDF  $p_i(s)$  of the  $i$ th extracted KIC is estimated, the probability of the  $i$ th KIC falling into the interval  $[s_{i,t} - \delta/2, s_{i,t} + \delta/2)$  can be calculated according to

$$\begin{aligned} f_i(s_{i,t}) &= \int_{s_{i,t} - \delta/2}^{s_{i,t} + \delta/2} p_i(s) ds = \xi_i \left( \int_{-\infty}^{\frac{s_{i,t} + \delta/2}{\sigma_{i,1}}} g_0(x; 1) dx \right. \\ &\quad \left. - \int_{-\infty}^{\frac{s_{i,t} - \delta/2}{\sigma_{i,1}}} g_0(x; 1) dx \right) + (1 - \xi_i) \left( \int_{-\infty}^{\frac{s_{i,t} + \delta/2}{\sigma_{i,2}}} g_0(x; 1) dx \right. \\ &\quad \left. - \int_{-\infty}^{\frac{s_{i,t} - \delta/2}{\sigma_{i,2}}} g_0(x; 1) dx \right), \end{aligned} \quad (23)$$

where  $\delta$  is a small positive constant which may be set to 0.1 without loss of generality, and the integral  $\int_{-\infty}^{\bar{x}} g_0(x; 1) dx$  is the value of the zero-mean unit-variance Gaussian cumulative distribution function (CDF) at the point  $\bar{x}$ . The look-up table of the standard Gaussian CDF can be stored, and the task of on-line computing  $f_i(s_{i,t})$  becomes the one that simply retrieves the four points of the standard Gaussian CDF from the look-up table. The value of  $f_i(s_{i,t})$  is a quantitative measure of the process operating state associated with  $s_i$  at the time  $t$ . As the correlation usually exists in the samples of  $s_i$ , the mean of the probabilities  $f_i(s_{i,t}), f_i(s_{i,t-1}), \dots, f_i(s_{i,t-q+1})$  over  $q$  samples can be calculated according to

$$\bar{f}_{i,t} = \frac{1}{q} \sum_{j=0}^{q-1} f_i(s_{i,t-j}), \quad (24)$$

to better capture the process operating information associated with  $s_i$ . Hence,  $\bar{f}_{i,t}$  measures the importance of the KIC  $s_i$  at the time  $t$  in revealing the useful fault information.

According to  $\bar{f}_{i,t}$ , the following rule is used to determine the weights in (15) and (16) at time  $t$

$$w_{i,t} = \begin{cases} \eta, & \text{if } \bar{f}_{i,t} > \bar{f}_{i,\text{lim}}, \\ 1 - \eta, & \text{if } \bar{f}_{i,t} \leq \bar{f}_{i,\text{lim}}, \end{cases} \quad (25)$$

where  $0 < \eta < 0.5$  is a pre-specified value, and  $\bar{f}_{i,\text{lim}}$  is a probability threshold for determining whether the KIC  $s_i$  is important for reflecting the useful process information at the time  $t$ . If  $\bar{f}_{i,t} > \bar{f}_{i,\text{lim}}$ , the sample  $s_{i,t}$  is considered to be within the normal operating region and contains no fault information. Then, a smaller weight value  $\eta$  is given to the sample  $s_{i,t}$  to suppress the irrelevant information provided by  $s_{i,t}$ . On the other hand, the condition  $\bar{f}_{i,t} \leq \bar{f}_{i,\text{lim}}$  suggests that the sample  $s_{i,t}$  may be abnormal and contains fault information. Thus, a larger weight value  $1 - \eta$  is assigned to  $s_{i,t}$  to highlight the significant fault information.

An empirical method for choosing the value of  $\eta$  is provided here. As the false alarm rate, defined as the percentage of the false alarming samples in all the normal operation samples, is

vital for measuring the reliability of fault detection [4], [5], [8], [14], it is taken as the evaluation index for choosing an appropriate value of  $\eta$ . Specifically, it is suggested to set the value of  $\eta$  to 0.3 initially and to check whether the false alarm rate of the normal-operation validating data is in the acceptable confidence range. If the false alarm rate is exceeding the confidence limit, the value of  $\eta$  is increased or reduced by  $\Delta\eta$  sequentially in the range of  $0 < \eta < 0.5$  until a modest false alarm rate is achieved.

The probability threshold  $\bar{f}_{i,\text{lim}}$  is another important parameter that must be chosen appropriately. As the samples  $\{s_{i,t}\}_{t=1}^n$  under the normal operating condition have affluent information for describing the normal operating region associated with the  $i$ th KIC  $s_i$ , the corresponding probabilities  $\{\bar{f}_{i,t}\}_{t=q}^n$  can be used to help choosing an appropriate value for  $\bar{f}_{i,\text{lim}}$ . Specifically, for the given  $\alpha$  confidence limit,  $(n - q + 1)(1 - \alpha)$  is rounded towards the nearest integer, denoted as  $r$ , and then the  $r$ th lowest value of  $\{\bar{f}_{i,t}\}_{t=q}^n$  is chosen as the probability threshold  $\bar{f}_{i,\text{lim}}$ .

In order to apply  $WI_t^2$  and  $WQ_t$  for determining whether the process is in the normal operating state or not, i.e., for fault detection, the corresponding confidence limits must be set. As no prior knowledge is available regarding the distributions of the KICs extracted by KICA in the built monitoring statistics, both the  $\alpha$  confidence limit  $WI_{\text{lim},\alpha}^2$  for the monitoring statistic  $WI_t^2$  and the  $\alpha$  confidence limit  $WQ_{\text{lim},\alpha}$  for the monitoring statistic  $WQ_t$  are determined by the well-known kernel density estimation (KDE) method [4], [5], [13], [14].

The proposed fault detection strategy using the GMM based WKICA includes the off-line modeling stage and the on-line fault detection stage, which are now summarized as follows.

*The off-line modeling stage:*

- 1) Collect the data from the process under normal operating conditions, and divide the process data into the training dataset and the validating dataset.
- 2) Based on the training data, construct the KICA model to obtain the training samples of the KICs.
- 3) Fit the probability density function of each KIC by using the GMM (17) and estimate the GMM parameters based on the training samples of each KIC using (19) to (22).
- 4) Calculate the probabilities of the training samples of each KIC using (23) and (24), and determine the probability threshold for each KIC.
- 5) Set the value of  $\eta$  in (25) to 0.3 initially.
- 6) Obtain the weight values for the training samples of each KIC using (25).
- 7) Use the obtained weight values in (15) and (16) to calculate the two monitoring statistics for the training data and determine the corresponding  $\alpha$  confidence limits using the KDE method.
- 8) Based on the validating data, use the built KICA model to obtain the validating samples of the KICs.
- 9) Calculate the probabilities of the validating samples of each KIC using (23) and (24).
- 10) Obtain the weight values for the validating samples of each KIC using (25).
- 11) Use the obtained weight values in (15) and (16) to calculate the two monitoring statistics for the validating

data. Compare the calculated monitoring statistics with their corresponding confidence limits to obtain their associated false alarm rates.

- 12) For each monitoring statistic, if the corresponding false alarm rate exceeds the pre-specified reasonable confidence range, the value of  $\eta$  is increased or reduced by  $\Delta\eta$  within the range of  $0 < \eta < 0.5$ , and go back to Step 6); Otherwise the modeling stage is completed.

*The on-line fault detection stage:*

- 1) Take the current data measurement from the process under monitoring.
- 2) Use the built KICA model to obtain the current samples of the KICs.
- 3) Calculate the probability of the current sample of each KIC using (23) and (24).
- 4) Obtain the weight value for the current sample of each KIC using (25).
- 5) Use the current weight values in (15) and (16) to calculate the current two monitoring statistics.
- 6) For each current monitoring statistic, compare it with its confidence limit. If the confidence limit is exceeded, proceed to the next step; Otherwise, go back to Step 1) for the next measurement.
- 7) Check whether the confidence limit is exceeded consecutively for the predefined number of samples. If yes, an abnormal process behavior is detected and an alarm should be given; Otherwise, go back to Step 1) for the next measurement.

In the above procedure, the training data, the validating data and the current monitoring process data are all normalized with the means and variances of the process variables calculated using the training data.

*Remarks:* Currently, there exist two fault detection methods [17], [31] using the KDE based weighted KPCA (WKPCA) and the KDE based weighted ICA (WICA), respectively. In the WKPCA, the weight allocation is conducted according to the importance of each KPC evaluated according to the PDF of the corresponding KPC which is estimated using KDE. Similarly, the weight allocation in the WICA is conducted according to the importance of each IC evaluated by the PDF of the corresponding IC, also estimated by KDE. However, the WKPCA is only applicable for handling Gaussian data with nonlinear structure while the WICA is only suitable for dealing with non-Gaussian data with linear structure. By contrast, our proposed WKICA is capable of effectively dealing with nonlinear and non-Gaussian data which are more common in real-world industrial environments and more challenging.

Both the WKPCA and WICA adopt KDE to estimate PDF. KDE uses all the  $n$  training samples to estimate the PDF of a new sample on-line. Generally, the training data size  $n$  should be sufficiently large in order to guarantee the accuracy of density estimation. Thus the complexity of on-line calculating the corresponding probability weight values based on the KDE approach may become unacceptably high. Various sparse KDE (SKDE) methods [32]–[34] may be adopted instead to provide sparse estimates of PDF in order to alleviate the high on-line computation burden. From this viewpoint, our GMM based

PDF estimation, which uses the two Gaussian mixture to estimate PDF, minimizes the on-line computational complexity in real-time fault detection application, while having a desired capability of modeling non-Gaussian PDF.

TABLE I  
COMPARISON OF THE ON-LINE COMPUTATIONAL LOADS.

method	KICA-based		WKICA-based	
task	$I_t^2$	$Q_t$	$WI_t^2$	$WQ_t$
multipliers	$da + d$	$(a - d)(a + 1)$	$da + 2d$	$(a - d)(a + 2)$
adders	$da - 1$	$(a - d)a - 1$	$da - 1$	$(a - d)a - 1$

*On-line fault detection complexity:* The on-line computational loads required by the KICA-based and WKICA-based methods to calculate their corresponding monitoring statistics are summarized in Table I, where it is seen that the on-line computational load for  $WI_t^2$  and  $WQ_t$  is slightly higher ( $a$  multiplications more) than that of  $I_t^2$  and  $Q_t$ . For the WKICA-based method, additionally the weights used in  $WI_t^2$  and  $WQ_t$  need to be determined on-line based on (23). Since the values of standard Gaussian CDF are calculated off line and stored in the memory space, the on-line computing of the probabilities  $f_i(s_{i,t})$  for  $1 \leq i \leq a$  becomes retrieving the values from the memory, which is extremely fast.

#### B. Fault identification based on nonlinear contribution-plots

After a fault is detected, it is the task of fault identification to identify fault variables in order to find the root cause of the fault. This task is very challenging, and few existing works discussed the methods of identifying faults. In particular, there is hardly any fault identification research for KICA-based monitoring methods owing to the usage of implicit nonlinear transformation. Based on the idea of sensitivity analysis [25], we develop a novel nonlinear contribution plots method for fault identification. Specifically, define

$$\mathbf{C}_{WI_t^2} = \frac{\partial WI_t^2}{\partial \mathbf{x}_t} \circ \mathbf{x}_t, \quad (26)$$

$$\mathbf{C}_{WQ_t} = \frac{\partial WQ_t}{\partial \mathbf{x}_t} \circ \mathbf{x}_t, \quad (27)$$

where  $\mathbf{C}_{WI_t^2} \in \mathbb{R}^m$  and  $\mathbf{C}_{WQ_t} \in \mathbb{R}^m$  represent the process variables' contributions to  $WI_t^2$  and  $WQ_t$ , respectively, while  $\circ$  denotes the Hadamard product. The explicit expressions of  $\mathbf{C}_{WI_t^2}$  and  $\mathbf{C}_{WQ_t}$  are given in Appendix.

Firstly, it is the absolute value of each process variable's contribution that reflects the influence of a process variable to  $WI_t^2$  and  $WQ_t$  [10], while its sign is unimportant. Secondly, the contributions of different process variables have different means and variances. Therefore, normalization is necessary to achieve accurate and reliable fault identification results. We suggest to use the mean and variance of each process variable's contribution sequence under the normal operating condition to scale the current contribution of each process variable in on-line operation. Thirdly, process variables' contributions are frequently affected by many uncertainties such as disturbances or measurement noises. The average contribution over a period of  $L$  fault samples provides a more robust and accurate indication. Based on the above analysis, the fault identification procedure using the proposed nonlinear contribution plots method can now be summarized as follows.

- 1) Calculate the process variables' contribution series  $\{\mathbf{C}_{WI_t^2}\}_{t=1}^n$  and  $\{\mathbf{C}_{WQ_t}\}_{t=1}^n$  based on the normal operating data, and compute the mean and variance of each process variable's contribution.
- 2) Suppose that a fault is detected by the monitoring statistics  $WI_t^2$  and  $WQ_t$  at sample times  $t_1$  and  $t_2$ , respectively. Let  $t_{\min} = \min\{t_1, t_2\}$ .
- 3) Calculate the process variables' contribution series  $\{\mathbf{C}_{WI_t^2}\}_{t=t_{\min}}^{t_{\min}+L-1}$  and  $\{\mathbf{C}_{WQ_t}\}_{t=t_{\min}}^{t_{\min}+L-1}$ , and scale them with the means and variances obtained in Step 1). Denote the absolute values of the normalized contribution series by  $\{|\bar{\mathbf{C}}_{WI_t^2}|\}_{t=t_{\min}}^{t_{\min}+L-1}$  and  $\{|\bar{\mathbf{C}}_{WQ_t}|\}_{t=t_{\min}}^{t_{\min}+L-1}$ .
- 4) The means of  $\{|\bar{\mathbf{C}}_{WI_t^2}|\}_{t=t_{\min}}^{t_{\min}+L-1}$  provide the contributions of the process variables to the monitoring statistic  $WI_t^2$ , while the means of  $\{|\bar{\mathbf{C}}_{WQ_t}|\}_{t=t_{\min}}^{t_{\min}+L-1}$  provide the contributions of the process variables to the monitoring statistic  $WQ_t$ .

## IV. SIMULATION STUDIES

The performance of the proposed GMM based WKICA monitoring scheme was compared with that of the conventional KICA method in the two case studies.

#### A. A four-variable system

A nonlinear and non-Gaussian process was simulated by the following four-variable nonlinear system

$$\begin{cases} x_1 = \frac{1}{2}b, \\ x_2 = -2b^2 + 0.2, \\ x_3 = \frac{1}{5} \exp(b + 1) - 0.56, \\ x_4 = \frac{\ln(b^2 + 1)}{4 \ln(2)} + \frac{1}{2}b, \end{cases} \quad (28)$$

where  $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]^T \in \mathbb{R}^4$  contained the four nonlinear and non-Gaussian output variables measured for fault detection and fault identification, while  $b = 0.3b_0$  and  $b_0$  was the non-Gaussian input variable whose PDF was given by

$$p(b_0) = \xi g_0(b_0; \sigma_1^2) + (1 - \xi) g_0(b_0; \sigma_2^2), \quad (29)$$

with  $\xi = 0.2$ ,  $\sigma_1^2 = 0.09$  and  $\sigma_2^2 = 1.2275$ . 1500 samples collected under the normal operating condition were divided into the training set of 1000 samples and the validation set of 500 samples. The following two fault cases were investigated:

- Fault 1: A step change in the process variable  $x_4$  with the step changing value of -0.15.
- Fault 2: A ramp change in the process variable  $x_1$  with the ramp changing rate of 0.0005.

For each fault pattern, 1000 samples were generated with the fault introduced at the 101th sample.

The fault detection performance was measured by the fault detection time and fault detection rate. To decrease the risk of false alarm, a fault is detected only when eight consecutive monitoring statistic values exceed the confidence limit, and the fault detection time is then defined as the first sample at which the confidence limit is exceeded. The fault detection rate is defined as the ratio of the fault samples whose monitoring

TABLE II

FALSE ALARM RATES (%) ASSOCIATED WITH THE TWO MONITORING STATISTICS OF EACH METHOD FOR THE VALIDATION DATASET OF THE FOUR-VARIABLE SYSTEM, GIVEN THE 99% CONFIDENCE LIMIT.

KICA-based method		WKICA-based method	
$I_t^2$	$Q_t$	$WI_t^2$	$WQ_t$
0.40	0.40	0.60	0.60

statistic values exceed the confidence limit to all the fault samples [7], [10], [14], [16], [19]. The hyper-dimension  $a$ , the number of dominant KICs  $d$ , and the kernel width  $c$  were found according to the empirical rule suggested in [14]. Specifically,  $a$  was determined as the number of the eigenvalues of the matrix  $\bar{K}$  which satisfy the condition

$$\lambda_i / \sum_{j=1}^n \lambda_j > 0.0001, i \in \{1, 2, \dots, n\}. \quad (30)$$

For both the conventional KICA and proposed WKICA methods, by setting  $c$  to 8000,  $a$  became close to the number of the measured output variables  $m$ , and thus  $a = 4$  was chosen. The value of  $d = 2$  was determined according to the cut-off method using the average eigenvalue of  $\{\lambda_i\}_{i=1}^n$ . Thus, the first two KICs  $s_1$  and  $s_2$  were used in the  $I_t^2$  and  $WI_t^2$  monitoring statistics, while the other two KICs  $s_3$  and  $s_4$  were used in the  $Q_t$  and  $WQ_t$  monitoring statistics. The  $\alpha = 99\%$  confidence limit was adopted as the alarming threshold. For the both monitoring statistics  $WI_t^2$  and  $WQ_t$  of the WKICA-based method, the parameter  $\eta$  in (25) was set to 0.3, and the corresponding false alarm rates calculated based on the validation dataset of the normal operation<sup>1</sup> are shown in Table II. It can be seen that the false alarm rates of each monitoring method are all lower than 1% and thus are acceptable for the given 99% confidence limit. This suggests that  $\eta = 0.3$  is appropriate for both  $WI_t^2$  and  $WQ_t$  in this case, and there is no need to search for other values for  $\eta$ .

The fault detection results for the first fault pattern using the two monitoring methods are shown in Figs. 1 and 2, respectively, where the values of a monitoring statistic are plotted by solid line while the corresponding confidence limit is depicted by dash line. All the monitoring statistic values are normalized by the corresponding confidence limit. It can be seen from Fig. 1 (a) that after the occurrence of the fault at the 101th sample, the values of the  $I_t^2$  monitoring statistic are much lower than the corresponding confidence limit for most of the fault samples, and the  $I_t^2$  monitoring statistic fails to detect this fault with a fault detection rate of only 0.89%. By contrast, as shown in Fig. 2 (a), the  $WI_t^2$  monitoring statistic can effectively detect this fault at the 104th sample with a fault detection rate of 99.67%. Although both the  $Q_t$  and  $WQ_t$  monitoring charts can detect this fault immediately with 100% fault detection rate, the  $WQ_t$  monitoring statistic exceeds its confidence limit with much larger margin than the  $Q_t$  monitoring statistic, indicating that the  $WQ_t$  monitoring statistic is much more reliable in detecting this fault.

The fault detection performances of the two monitoring methods under the fault pattern 2 are shown in Figs. 3 and 4,

<sup>1</sup>Under the normal operating case, there is no fault occurring. The WKICA method will not find any KIC associated with fault information and consequently, it will automatically assign the equal low weight  $\eta$  to all the extracted KICs, i.e., it will use an equal weighting strategy similar to the KICA method.

respectively. Clearly, the WKICA method has a superior fault detection performance over the KICA method for this fault

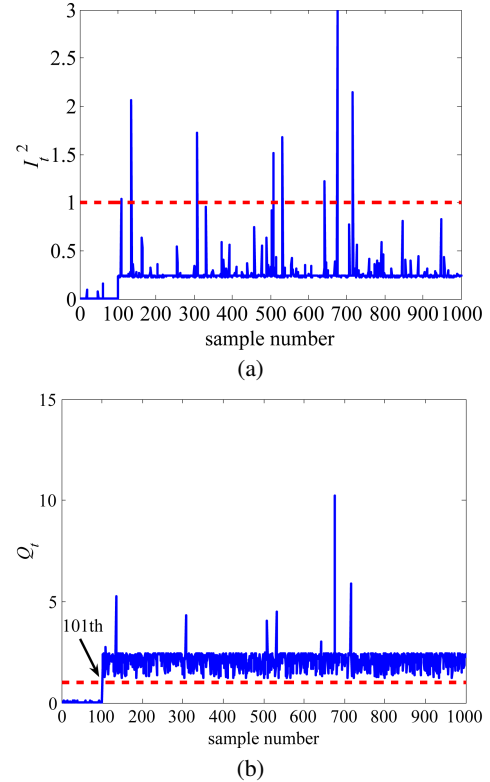


Fig. 1. (a)  $I_t^2$  monitoring chart and (b)  $Q_t$  monitoring chart of the KICA method for the four-variable system under the fault pattern 1. Dashed line indicates 99% confidence limit for the corresponding monitoring statistic.

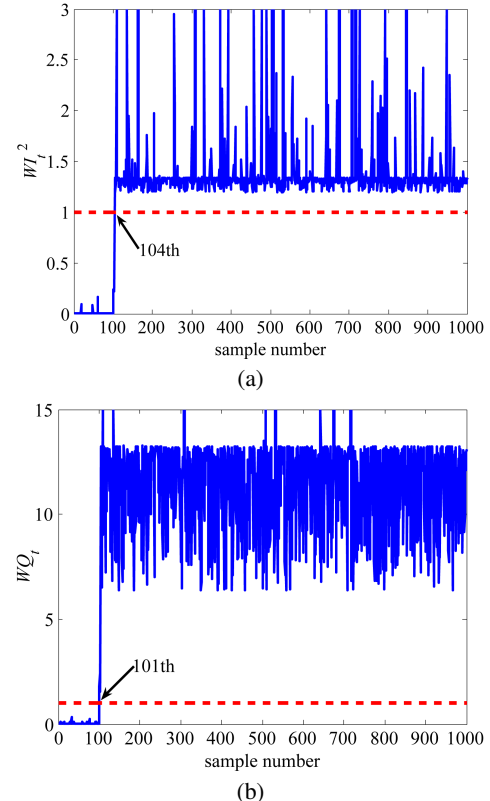


Fig. 2. (a)  $WI_t^2$  monitoring chart and (b)  $WQ_t$  monitoring chart of the WKICA method for the four-variable system under the fault pattern 1. Dashed line indicates 99% confidence limit for the corresponding monitoring statistic.

pattern. Specifically, the two monitoring charts of the WKICA method are able to detect the occurrence of the fault much ear-

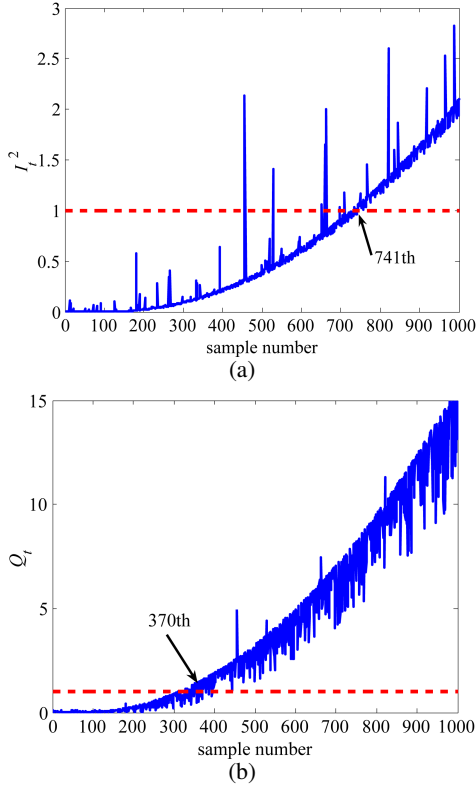


Fig. 3. (a)  $I_t^2$  monitoring chart and (b)  $Q_t$  monitoring chart of the KICA method for the four-variable system under the fault pattern 2. Dashed line indicates 99% confidence limit for the corresponding monitoring statistic.

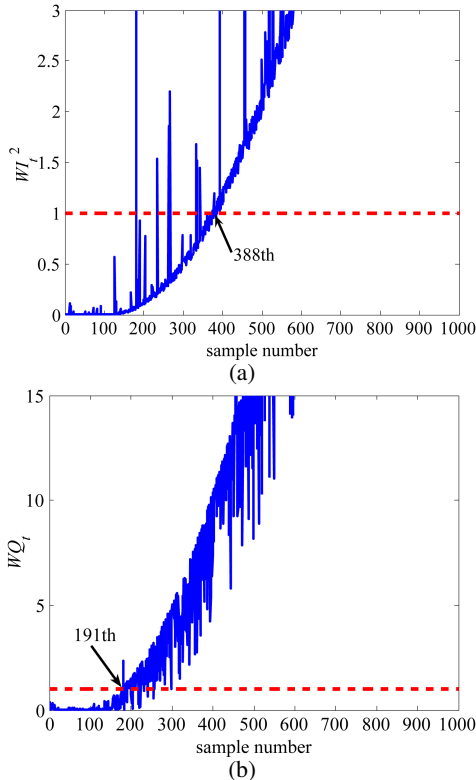


Fig. 4. (a)  $WI_t^2$  monitoring chart and (b)  $WQ_t$  monitoring chart of the WKICA method for the four-variable system under the fault pattern 2. Dashed line indicates 99% confidence limit for the corresponding monitoring statistic.

lier than the two corresponding monitoring charts of the KICA method. Moreover, the  $WI_t^2$  monitoring chart attains 70.00% fault detection rate, compared to 30.00% fault detection rate of the  $I_t^2$  monitoring chart, while the  $WQ_t$  monitoring chart achieves 89.67% fault detection rate, in comparison to 72.89% fault detection rate of the  $Q_t$  monitoring chart.

Fig. 5 plots the estimated probabilities  $\bar{f}_{i,t}$ , averaged over  $q = 8$  samples, of the KICs' samples  $s_{i,t}$  for  $1 \leq i \leq 4$ , where

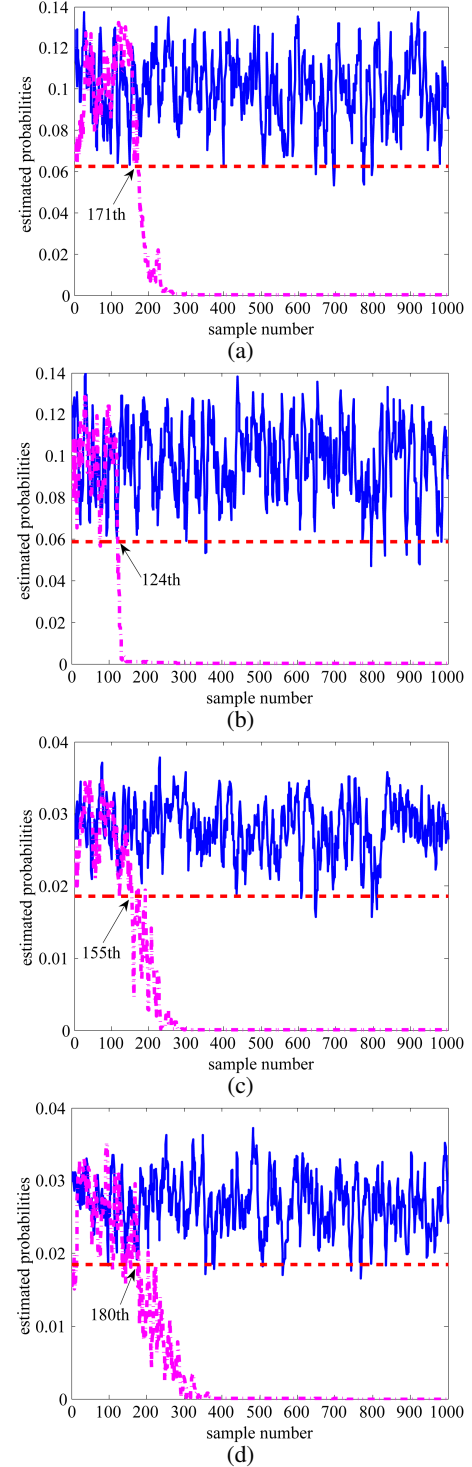


Fig. 5. Estimated  $\bar{f}_{i,t}$  for the samples of  $s_i$ : (a)  $i = 1$ , (b)  $i = 2$ , (c)  $i = 3$ , and (d)  $i = 4$ , for the training data (blue solid curve) and the data of the fault pattern 2 (magenta dot dash curve). Dashed line indicates probability threshold for the corresponding KIC.



the blue solid curve denotes the estimated probabilities of the KICs' samples from the training data, and the magenta dot dash curve represents the estimated probabilities of the KICs' samples from the fault-pattern-2 data, while the red dash line is the probability threshold  $\bar{f}_{i,\text{lim}}$  for the corresponding KIC. From Fig. 5, it can be seen that almost all the estimated probabilities of the KICs' samples for the training dataset are above the corresponding probability thresholds, indicating that the KICs' samples are not associated with a fault, and therefore the corresponding weights for the KICs' samples in constructing

the two monitoring charts take a lower value of  $\eta$  according to (25). By contrast, the magenta dot dash curves in Fig. 5 indicate that the estimated probabilities of the different KICs' samples fall below the corresponding probability thresholds after certain different sample numbers, which are the times that the four KICs can effectively reveal the fault information. Specifically, in the monitoring statistic  $WI_t^2$ , the dominant KICs  $s_1$  and  $s_2$  can reflect the occurring fault information from the 171th and 124th samples, respectively, while in the monitoring statistic  $WQ_t$ , the excluded KICs  $s_3$  and  $s_4$  can reveal the occurring fault information from the 155th and 180th samples, respectively. Accordingly, a higher weight  $1-\eta$  is assigned to the corresponding KICs' samples according to their online estimated probabilities.

After a fault is detected, it is important to identify the fault variables that cause the abnormal condition. The nonlinear fault identification is an unsolved open problem, especially for the nonlinear ICA based monitoring methods with kernel technique. Here we demonstrate the potential of our proposed nonlinear contribution plots method in fault identification. For the fault pattern 1, the  $WI_t^2$  and  $WQ_t$  monitoring charts detected the fault at the 104th and 101th samples, respectively. Accordingly, we set  $t_{\min} = 101$  and averaged the contribution values over  $L = 2$  fault samples. The fault identification results using the proposed contribution plots method are shown in Fig. 6, where it can be clearly seen that the fourth process variable has the largest contribution values to both  $WI_t^2$  and  $WQ_t$ . This indicates that  $x_4$  is the most likely process variable that causes the occurring fault. With this vital information, an engineer with the knowledge of the plant may then be able to identify the root cause of the occurring fault.

For the fault pattern 2, the  $WI_t^2$  and  $WQ_t$  monitoring statistics indicated a fault at the 388th and 191th samples, respectively. Thus, we set  $t_{\min} = 191$ , and again chose  $L = 2$ . The corresponding fault identification results obtained by our contribution plots method are demonstrated in Fig. 7. It can be clearly seen that the largest contributions in the both contribution plots come from the first process variable, and therefore  $x_1$  is correctly located as the fault variable.

### B. The Tennessee Eastman industrial process

The Tennessee Eastman (TE) industrial process [35] is a well-known benchmark process for testing process monitoring methods [1], [5], [7]–[10], [14], [17], [19], [20], [22]. The flowchart of the TE process is depicted in Fig. 8, which consists of five major units, a reactor, a stripper, a condenser, a compressor and a separator. There are totally 52 measured process variables for monitoring this process, including 11 manipulated variables (MVs), 22 continuous process measurements, and 19 composition measurements. All the measured process variables are listed in Table III. A simulator for the TE process coded by FORTRAN is provided by <http://brahms.scs.uiuc.edu>. It allows 21 pre-programmed process faults, which are listed in Table IV. Each fault pattern contains 960 samples with a sampling interval of 3 minutes and the corresponding fault is introduced at the 160th sample. The TE simulator also produces two normal operating datasets with 480 samples and 960 samples, respectively.

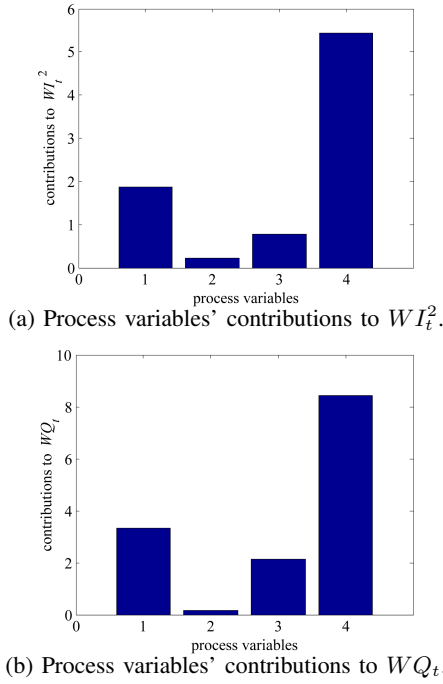


Fig. 6. The fault identification results using our nonlinear contribution plots method for the four-variable system under the fault pattern 1.

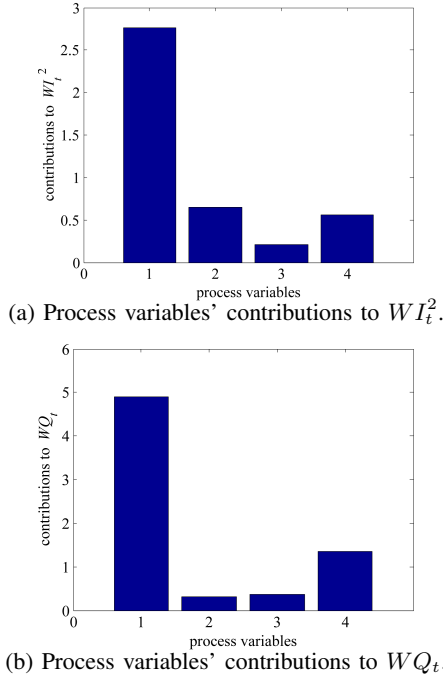


Fig. 7. The fault identification results using our nonlinear contribution plots method for the four-variable system under the fault pattern 2.

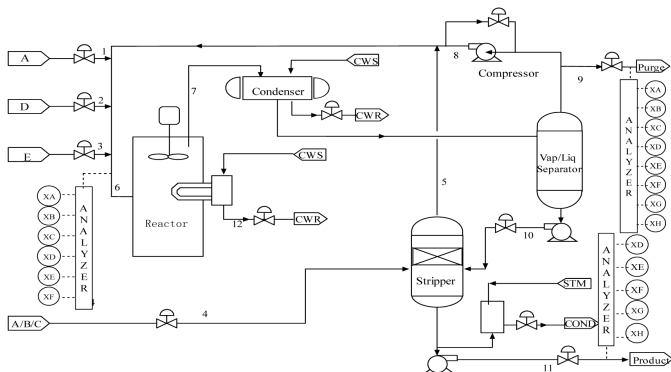


Fig. 8. The flowchart of the Tennessee Eastman industrial process.

The normal operating dataset with 960 samples was used for training, while the other normal operating dataset with 480 samples was adopted as the validating set for testing the false alarm rate of the built model. The hyper-dimension  $a$ , the number of the dominant KICs  $d$ , and the kernel width parameter  $c$  were determined based on the same empirical method used for the first case study, yielding  $c = 6000$ ,  $a = 54$  and  $d = 42$ . Thus, the first 42 KICs  $s_i$ ,  $1 \leq i \leq 42$ , were used in constructing the  $I_t^2$  and  $WI_t^2$  monitoring charts, while the last 12 KICs  $s_i$ ,  $43 \leq i \leq 52$ , were used in computing the  $Q_t$  and  $WQ_t$  monitoring statistics. Again the  $\alpha = 99\%$  confidence limit was adopted as the alarm threshold. For the both  $WI_t^2$  and  $WQ_t$  monitoring statistics, the parameter  $\eta$  was set to 0.3, and the estimated probabilities were averaged

TABLE III  
MEASURED PROCESS VARIABLES OF THE TENNESSEE EASTMAN PROCESS.

No.	Process variable
1	A feed (stream 1)
2	D feed (stream 2)
3	E feed (stream 3)
4	Total feed (Stream 4)
5	Recycle flow (Stream 8)
6	Reactor feed rate (Stream 6)
7	Reactor pressure
8	Reactor level
9	Reactor temperature
10	Purge rate (Stream 9)
11	Product separator temperature
12	Product separator level
13	Product separator pressure
14	Product separator underflow (Stream 10)
15	Stripper level
16	Stripper pressure
17	Stripper underflow (Stream 11)
18	Stripper temperature
19	Stripper steam flow
20	Compressor work
21	Reactor cooling water outlet temperature
22	Separator cooling water outlet temperature
23–28	Components A, B, C, D, E, F in stream 6
29–36	Components A, B, C, D, E, F, G, H in stream 9
37–41	Components D, E, F, G, H in stream 11
42	MV for D feed flow (stream 2)
43	MV for E feed flow (stream 3)
44	MV for A feed flow (stream 1)
45	MV for total feed flow (stream 4)
46	MV for compressor recycle valve
47	MV for purge valve (stream 9)
48	MV for separator pot liquid flow (stream 10)
49	MV for stripper liquid prod flow (stream 11)
50	MV for stripper steam valve
51	MV for reactor cooling water flow
52	MV for condenser cooling water flow

TABLE IV  
FAULT PATTERNS OF THE TENNESSEE EASTMAN PROCESS.

No.	Description	Fault type
1	A/C feed ratio, B composition constant (stream 4)	Step
2	B composition, A/C feed ratio constant (stream 4)	Step
3	D feed temperature (stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss (stream 1)	Step
7	C header pressure loss-reduced availability (stream 4)	Step
8	A, B, C feed compositions (stream 4)	Random variation
9	D feed temperature (stream 2)	Random variation
10	C feed temperature (stream 4)	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16–20	Unknown	Unknown
21	Valve position constant (stream 4)	Constant position

over  $q = 8$  KICs' samples. The false alarm rates of  $WI_t^2$  and  $WQ_t$  calculated based on the validating data are listed in Table V, together with the false alarm rates of the KICA method. Given the 99% confidence limit, the false alarm rates of the both monitoring methods can be regarded as within the acceptable confidence range. This suggested that  $\eta = 0.3$  is appropriate for both  $WI_t^2$  and  $WQ_t$ .

The monitoring results under the fault pattern 4 are given in Figs. 9 and 10, respectively, for the conventional KICA method and the proposed WKICA method. Comparing Fig. 9 (a) to Fig. 10 (a), it can be seen that after the fault occurred at the 160th sample, the  $I_t^2$  monitoring statistic's values do increase but fluctuate around the confidence limit, while almost all the  $WI_t^2$  monitoring statistic's values increase to well above the confidence limit. Thus, the  $WI_t^2$  monitoring chart can detect this fault much more confidently and reliably than the  $I_t^2$  monitoring chart. Similarly, observe from Fig. 9 (b) and Fig. 10 (b) that, although both the  $Q_t$  and  $WQ_t$  monitoring charts can detect the fault at the 163th sample with fault detection rates close to 100%, the  $WQ_t$  monitoring statistic exceeds its confidence limit with much larger margin than the  $Q_t$  monitoring statistic, indicating that the  $WQ_t$  monitoring statistic provides a more reliable detection of this fault.

The fault detection performances of our WKICA method are compared with those of the KICA method for the 18 representative fault patterns of the TE process in Tables VI and VII, respectively, in terms of fault detection time and fault detection rate. The fault patterns 3, 9 and 15 have been testified to be extremely difficult for the data-driven monitoring methods due to the reason that there are no observable changes in the mean or the variance of these fault datasets [35]. Therefore, we excluded these three fault patterns in our investigation. From Table VI, it is seen that the both methods achieve similar

TABLE V  
FALSE ALARM RATES (%) ASSOCIATED WITH THE TWO MONITORING STATISTICS OF EACH METHOD FOR THE TENNESSEE EASTMAN PROCESS, GIVEN THE 99% CONFIDENCE LIMIT.

KICA-based method		WKICA-based method	
$I_t^2$	$Q_t$	$WI_t^2$	$WQ_t$
0.00	1.87	0.00	1.67

fault detection times for most of the 18 fault patterns, but there are three fault patterns for which the monitoring charts

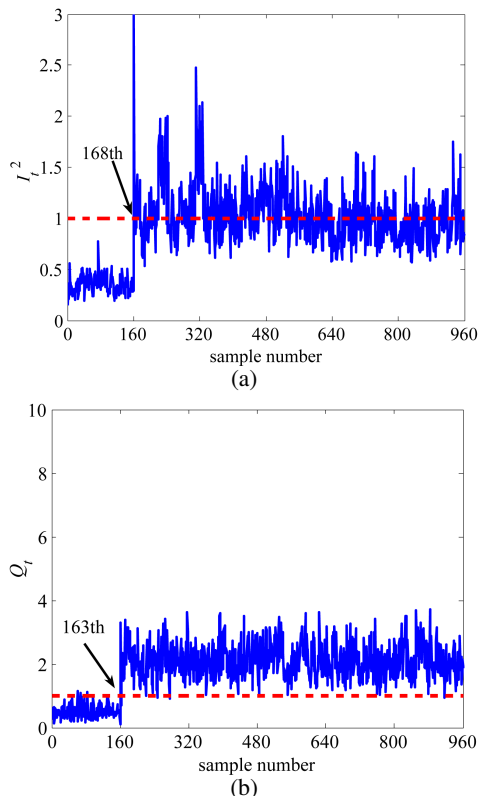


Fig. 9. (a)  $I_t^2$  monitoring chart and (b)  $Q_t$  monitoring chart of the KICA method for the TE process under the fault pattern 4. Dashed line indicates 99% confidence limit for the corresponding monitoring statistic.

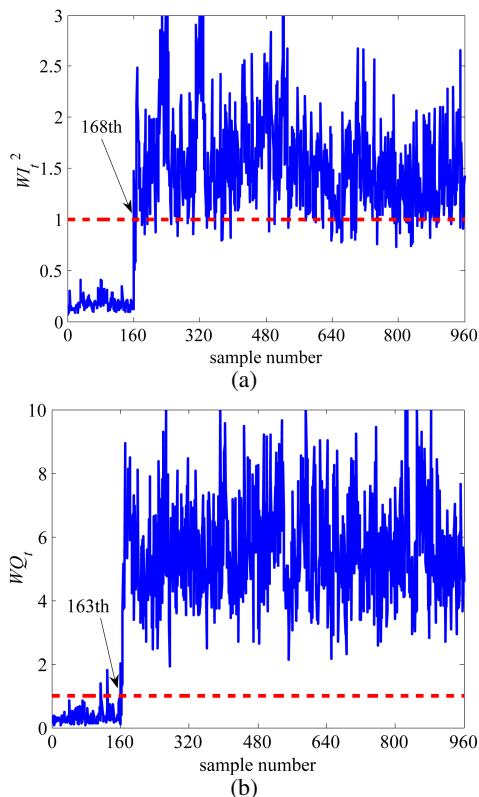


Fig. 10. (a)  $WI_t^2$  monitoring chart and (b)  $WQ_t$  monitoring chart of the WKICA method for the TE process under the fault pattern 4. Dashed line indicates 99% confidence limit for the corresponding monitoring statistic.

of our WKICA method detect the occurrence of fault much earlier than the monitoring charts of the KICA method. Fault detection time does not tell how reliable the detection is, and for measuring the reliability of fault detection, we need to turn to fault detection rate. Observe from Table VII that the WKICA method achieves better fault detection rates than the KICA method for one third of the fault patterns, while for the other two thirds of the fault patterns the both methods attain the similar performance. It can be seen that the proposed WKICA-based method is more powerful than the KICA-based method, particularly for detecting complex faults, including the fault patterns 10, 11, 16, 17, 19, 20 and 21.

The potential of the proposed nonlinear contribution plots method for fault identification was then demonstrated using the fault patterns 6 and 11. For the fault pattern 6, both  $WI_t^2$  and  $WQ_t$  detect the fault at the 161th sample. Hence, we set  $t_{\min} = 161$ , while  $L$  was set to 2. Fig. 11 shows the fault identification results for the fault pattern 6, where it can be clear seen that the 1st and 44th process variables have

TABLE VI  
FAULT DETECTION TIMES (SAMPLE NUMBER) ACHIEVED BY THE KICA-BASED AND WKICA-BASED METHODS FOR 18 FAULT PATTERNS OF THE TENNESSEE EASTMAN PROCESS.

Fault No.	KICA-based		WKICA-based	
	$I_t^2$	$Q_t$	$WI_t^2$	$WQ_t$
1	163	163	166	165
2	175	183	175	178
4	168	163	168	163
5	161	161	163	162
6	161	165	161	161
7	161	161	161	161
8	180	182	182	180
10	187	189	185	189
11	<b>256</b>	170	<b>171</b>	171
12	163	166	163	163
13	205	210	203	208
14	162	162	163	162
16	171	171	173	171
17	186	182	185	182
18	245	244	244	242
19	<b>failed</b>	<b>195</b>	<b>333</b>	<b>171</b>
20	245	241	242	241
21	<b>675</b>	<b>711</b>	<b>662</b>	<b>665</b>

TABLE VII  
FAULT DETECTION RATES (%) ACHIEVED BY THE KICA-BASED AND WKICA-BASED METHODS FOR 18 FAULT PATTERNS OF THE TENNESSEE EASTMAN PROCESS.

Fault No.	KICA-based		WKICA-based	
	$I_t^2$	$Q_t$	$WI_t^2$	$WQ_t$
1	99.50	99.75	99.38	99.50
2	98.38	98.00	98.38	98.25
4	<b>49.13</b>	99.38	<b>91.62</b>	99.88
5	24.13	26.75	25.62	28.25
6	100.0	99.88	100.0	100.0
7	96.88	100.0	99.88	100.0
8	97.63	97.63	97.38	97.75
10	<b>72.63</b>	<b>76.38</b>	<b>79.12</b>	<b>84.25</b>
11	<b>42.75</b>	<b>73.25</b>	<b>50.88</b>	<b>82.00</b>
12	99.13	97.63	99.25	98.75
13	94.50	94.75	94.75	95.25
14	99.88	99.88	99.75	99.88
16	<b>71.75</b>	<b>84.00</b>	<b>78.50</b>	<b>90.38</b>
17	<b>86.38</b>	96.13	<b>91.62</b>	97.25
18	89.63	90.38	89.62	90.50
19	<b>41.25</b>	<b>77.38</b>	<b>54.25</b>	<b>89.38</b>
20	<b>47.88</b>	<b>57.50</b>	<b>54.75</b>	<b>67.62</b>
21	<b>36.38</b>	<b>34.38</b>	<b>40.00</b>	<b>43.62</b>

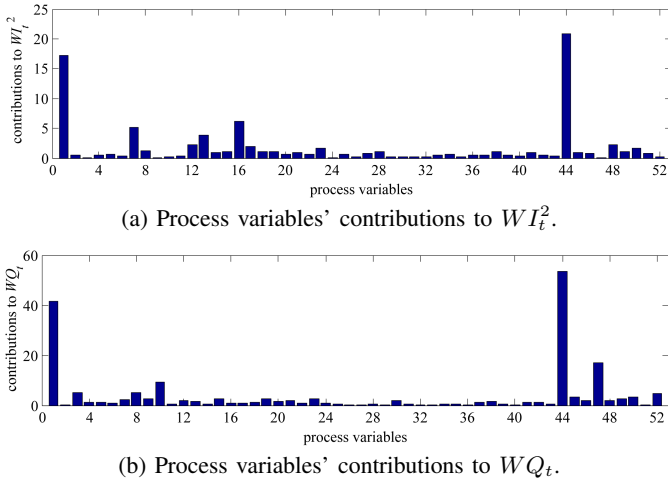


Fig. 11. The fault identification results using our nonlinear contribution plots method for the Tennessee Eastman process under the fault pattern 6.

the highest contributions to this fault. The fault pattern 6 is a step-type fault which is caused by the component  $A$  feed loss in the stream 1, as described in Table IV. Referring to the process variable description of the TE process given in Table III, the 1st variable is the component  $A$  feed in the stream 1 and the 44th variable is the component  $A$  feed flow valve in the stream 1. Thus, the contribution plots of Fig. 11 correctly identify the two process variables that are closely associated with the occurring fault and, therefore, they provide the direct and effective guidance for locating the fault source.

For the fault pattern 11, both the  $WI_t^2$  and  $WQ_t$  monitoring statistics detect the fault at the 171th sample. Thus, we set  $t_{\min} = 171$ , and chose again  $L = 2$ . The fault identification results are illustrated in Fig. 12, where it can be observed that the two largest contributions in the both contribution plots come from the 9th and 51th process variables. The 9th variable is the reactor temperature and the 51th variable is the reactor cooling water flow valve. Actually, the fault pattern 11 is caused by the random variation of the reactor cooling water inlet temperature, according to the fault description given in Table 3. However, the reactor cooling water inlet temperature is not contained in the set of the TE process's monitored

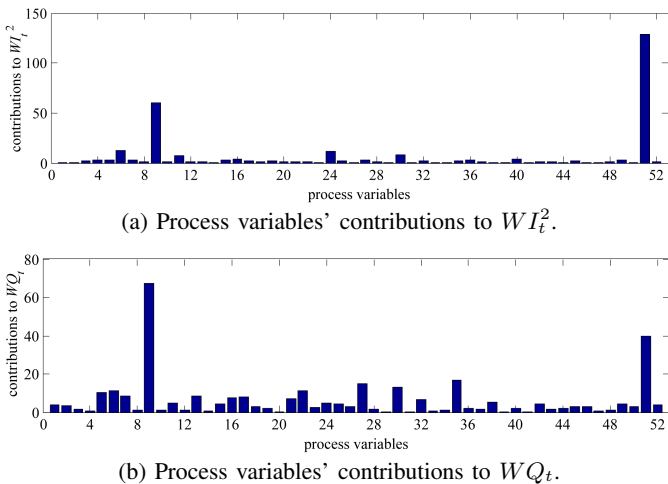


Fig. 12. The fault identification results using our nonlinear contribution plots method for the Tennessee Eastman process under the fault pattern 11.

variables. From the underlying TE process knowledge, it can be easily found that the reactor cooling water inlet temperature is closely connected with the reactor temperature and the reactor cooling water flow valve. Thus, by indicating that the 9th and 51th process variables are the main contributors to the occurring fault, the contribution plots of Fig. 12 provides the plant operator who has the underlying plant knowledge with the valuable information to locate the root of the fault.

## V. CONCLUSIONS

A monitoring method for nonlinear and non-Gaussian processes has been proposed using the novel GMM based WKICA approach. In our WKICA method, KICA is used to extract the KICs from the process data. However, unlike the existing KICA method, the PDFs of the extracted KICs are estimated using the two-Gaussian mixture, to provide the estimated probabilities of the KICs' samples online for highlighting significant fault information and suppressing the information irrelevant to the occurring fault. Simulation results obtained on the four-variable nonlinear and non-Gaussian system and the Tennessee Eastman benchmark process have demonstrated that the proposed GMM based WKICA method outperforms the existing state-of-the-art KICA method, in terms of false detection time and fault detection rate. We have also investigated the challenging problem of fault identification for kernel based methods, and have proposed a nonlinear contribution plots method. The potential of this contribution plots method in identifying the underlying fault variables has been demonstrated in the two case studies. As a concluding remark, we point out that the determination of the optimal kernel width for nonlinear process monitoring and fault identification are two open problems. Further works are warranted to provide theoretical direction and practical implementation.

## APPENDIX

Noting  $s_{d,t} = U_d \sqrt{n} \Lambda_a^{-1} H_a^T \bar{k}_t$ , the monitoring statistic  $WI_t^2$  can be expressed as

$$WI_t^2 = \bar{k}_t^T M_1 \bar{k}_t, \quad (31)$$

where

$$M_1 = n H_a \Lambda_a^{-1} U_d^T \text{diag}\{w_{1,t}^2, \dots, w_{d,t}^2\} U_d \Lambda_a^{-1} H_a^T. \quad (32)$$

Similarly, since  $s_{e,t} = U_e \sqrt{n} \Lambda_a^{-1} H_a^T \bar{k}_t$ , we have

$$WQ_t = \bar{k}_t^T M_2 \bar{k}_t, \quad (33)$$

where

$$M_2 = n H_a \Lambda_a^{-1} U_e^T \text{diag}\{w_{d+1,t}^2, \dots, w_{a,t}^2\} U_e \Lambda_a^{-1} H_a^T. \quad (34)$$

By substituting (31) into (26), we obtain

$$\begin{aligned} C_{WI_t^2} &= \left( \left( \left[ \frac{\partial WI_t^2}{\partial \bar{k}_t} \dots \frac{\partial WI_t^2}{\partial \bar{k}_t} \right]^T \circ \left[ \frac{\partial \bar{k}_t}{\partial x_{1,t}} \dots \frac{\partial \bar{k}_t}{\partial x_{m,t}} \right]^T \right) \right. \\ &\quad \times \text{one}_n \left. \right) \circ x_t \\ &= \left( \left( \left[ 2M_1 \bar{k}_t \dots 2M_1 \bar{k}_t \right]^T \circ \left[ \frac{\partial \bar{k}_t}{\partial x_{1,t}} \dots \frac{\partial \bar{k}_t}{\partial x_{m,t}} \right]^T \right) \right. \\ &\quad \times \text{one}_n \left. \right) \circ x_t, \end{aligned} \quad (35)$$

where  $\mathbf{one}_n$  denotes the  $n \times 1$  column vector whose elements are all equal to one, while  $\mathbf{x}_t = [x_{1,t} \ x_{2,t} \ \cdots \ x_{m,t}]^T$  is the current process variable vector.

Similarly, substituting (33) into (27) leads to

$$C_{WQ_t} = \left( \left( [2M_2 \bar{\mathbf{k}}_t \ \cdots \ 2M_2 \bar{\mathbf{k}}_t]^T \circ \left[ \frac{\partial \bar{\mathbf{k}}_t}{\partial x_{1,t}} \ \cdots \ \frac{\partial \bar{\mathbf{k}}_t}{\partial x_{m,t}} \right]^T \right) \times \mathbf{one}_n \right) \circ \mathbf{x}_t. \quad (36)$$

The derivatives  $\frac{\partial \bar{\mathbf{k}}_t}{\partial x_{j,t}}$ ,  $1 \leq j \leq m$ , are given by

$$\begin{aligned} \frac{\partial \bar{\mathbf{k}}_t}{\partial x_{j,t}} &= \frac{\partial (\mathbf{k}_t - \mathbf{K} \mathbf{1}_1 - \mathbf{1}_n \mathbf{k}_t + \mathbf{1}_n \mathbf{K} \mathbf{1}_1)}{\partial x_{j,t}} \\ &= (\mathbf{I}_n - \mathbf{1}_n) \frac{\partial \mathbf{k}_t}{\partial x_{j,t}}, \end{aligned} \quad (37)$$

in which

$$\begin{aligned} \frac{\partial \mathbf{k}_t}{\partial x_{j,t}} &= \left[ \exp \left( -\frac{\|\mathbf{x}_1 - \mathbf{x}_t\|^2}{c} \right) \left( \frac{-2}{c} \right) (x_{j,t} - x_{j,1}) \ \cdots \right. \\ &\quad \left. \exp \left( -\frac{\|\mathbf{x}_n - \mathbf{x}_t\|^2}{c} \right) \left( \frac{-2}{c} \right) (x_{j,t} - x_{j,n}) \right]^T. \end{aligned} \quad (38)$$

## REFERENCES

- [1] S. J. Qin and Y. Y. Zheng, "Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures," *AICHE Journal* vol. 59, no. 2 pp. 496–504, Feb. 2013.
- [2] X. M. Tian, L. F. Cai, and S. Chen, "Noise-resistant joint diagonalization independent component analysis based process fault detection," *Neurocomputing*, vol. 149, pp. 652–666, 2015.
- [3] J. S. Zeng, L. Xie, U. Kruger, and C. H. Gao, "Regression-based analysis of multivariate non-Gaussian datasets for diagnosing abnormal situations in chemical processes," *AICHE Journal*, vol. 60, no. 1, pp. 148–159, Jan. 2014.
- [4] Y. Zhang, J. An, and H. Zhang, "Monitoring of time-varying processes using kernel independent component analysis," *Chemical Engineering Science*, vol. 88, pp. 23–32, Jan. 2013.
- [5] Z. Ge and Z. Song, "Performance-driven ensemble learning ICA model for improved non-Gaussian process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 123, pp. 1–8, Apr. 2013.
- [6] C. Zhao and F. Gao, "Fault-relevant principal component analysis (FPCA) method for multivariate statistical modeling and process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 133, pp. 1–16, Apr. 2014.
- [7] S. Stubbs, J. Zhang, and J. Morris, "Fault detection in dynamic processes using a simplified monitoring-specific CVA state space modeling approach," *Computers and Chemical Engineering*, vol. 41, pp. 77–87, Jun. 2012.
- [8] H. Ma, Y. Hu, and H. Shi, "A novel local neighborhood standardization strategy and its application in fault detection of multimode processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 287–300, Aug. 2012.
- [9] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part I," *Signal Processing*, vol. 94, pp. 278–287, Jan. 2014.
- [10] 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, Aug. 2013.
- [11] J. Zhou, A. Guo, B. Celler, and S. Su, "Fault detection and identification spanning multiple processes by integrating PCA with neural network," *Applied Soft Computing*, vol. 14, Part A, pp. 4–11, Jan. 2014.
- [12] Y. Yang, X. Li, X. Liu, and X. Chen, "Wavelet kernel entropy component analysis with application to industrial process monitoring," *Neurocomputing*, vol. 147, pp. 395–402, Jan. 2015.
- [13] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AICHE Journal*, vol. 52, no. 10, pp. 3501–3514, Oct. 2006.
- [14] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection of non-linear processes using kernel independent component analysis," *The Canadian J. Chemical Engineering*, vol. 85, no. 4, pp. 526–536, Aug. 2007.
- [15] J. Yu, "A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes," *Chemical Engineering Science*, vol. 68, no. 1, pp. 506–519, Jan. 2012.
- [16] L. Cai, X. Tian, and S. Chen, "A process monitoring method based on noisy independent component analysis," *Neurocomputing*, vol. 127, pp. 231–246, Mar. 2014.
- [17] Q. Jiang, X. Yan, and C. Tong, "Double-weighted independent component analysis for non-Gaussian chemical process monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 40, pp. 14396–14405, Sep. 2013.
- [18] M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, "Monitoring independent components for fault detection," *AICHE Journal*, vol. 49, no. 4, pp. 969–976, Apr. 2003.
- [19] P. P. Odiowei and Y. Cao, "State-space independent component analysis for nonlinear dynamic process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 103, no. 1, pp. 59–65, Aug. 2010.
- [20] M. M. Rashid and J. Yu, "Hidden Markov model based adaptive independent component analysis approach for complex chemical process monitoring and fault detection," *Industrial & Engineering Chemistry Research*, vol. 51, no. 15, pp. 5506–5514, Mar. 2012.
- [21] J. Yang, X. Gao, D. Zhang, and J.-Y. Yang, "Kernel ICA: An alternative formulation and its application to face recognition," *Pattern Recognition*, vol. 38, no. 10, pp. 1784–1787, Oct. 2005.
- [22] J. Fan, S. J. Qin, and Y. Wang, "Online monitoring of nonlinear multivariate industrial processes using filtering KICA-PCA," *Control Engineering Practice*, vol. 22, pp. 205–216, Jan. 2014.
- [23] X. Tian, X. Zhang, X. Deng, and S. Chen, "Multiway kernel independent component analysis based on feature samples for batch process monitoring," *Neurocomputing*, vol. 72, nos. 7–9, pp. 1584–1596, Mar. 2009.
- [24] I. Santamaría, C. J. Pantaleón, J. Ibáñez, and A. Artés, "Deconvolution of seismic data using adaptive Gaussian mixtures," *IEEE Trans. Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 855–859, Mar. 1999.
- [25] L. Petzold, S. Li, Y. Cao, and R. Serban, "Sensitivity analysis of differential-algebraic equations and partial differential equations," *Computers & Chemical Engineering*, vol. 30, nos. 10–12, pp. 1553–1559, Sep. 2006.
- [26] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.
- [27] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
- [28] T. Chen and J. Zhang, "On-line multivariate statistical monitoring of batch processes using Gaussian mixture model," *Computers & Chemical Engineering*, vol. 34, no. 4, pp. 500–507, Apr. 2010.
- [29] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Computers & Chemical Engineering*, vol. 28, no. 8, pp. 1377–1387, Jul. 2004.
- [30] Y. Zhao, S. Zhuang, and S.-J. Ting, "Gaussian mixture density modeling of non-Gaussian source for autoregressive process," *IEEE Trans. Signal Processing*, vol. 43, no. 4, pp. 894–903, Apr. 1995.
- [31] Q. Jiang and X. Yan, "Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 127, pp. 121–131, Aug. 2013.
- [32] M. Girolami and C. He, "Probability density estimation from optimally condensed data samples," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1253–1264, Oct. 2003.
- [33] X. Hong, S. Chen, and C. J. Harris, "A forward-constrained regression algorithm for sparse kernel density estimation," *IEEE Trans. Neural Networks*, vol. 19, no. 1, pp. 193–198, Jan. 2008.
- [34] S. Chen, X. Hong, and C. J. Harris, "Particle swarm optimization aided orthogonal forward regression for unified data modelling," *IEEE Trans. Evolutionary Computation*, vol. 14, no. 4, pp. 477–499, Aug. 2010.
- [35] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. Springer-Verlag, London, 2001.



**Lianfang Cai** received his Bachelor of Engineering degree from China University of Petroleum (Hua Dong) in July 2009, and his PhD degree from China University of Petroleum (Hua Dong) in December 2014.

He is currently a postdoctoral researcher in Imperial College London, UK. His research interests are in data-driven fault detection and diagnosis, multivariate statistical analysis and blind signal processing.



**Xuemin Tian** received his Bachelor of Engineering degree from Huadong Petroleum Institute, Dongying, China, in January 1982, and his M.S. degree from Beijing University of Petroleum, Beijing, China, in June 1994. From September 2001 to June 2002, he served as visiting professor at central of process control, University of California in Santa Barbara.

He is a professor of Process Control at China University of Petroleum (Hua Dong). Professor Tian's research interests are in modelling, advanced process

control and optimisation for petrol-chemical processes as well as fault detection and diagnosis, and process monitoring.



**Sheng Chen** (M'1990-SM'1997-F'2008) received his BEng degree from the East China Petroleum Institute, China, in January 1982, and his PhD degree from the City University, London, in September 1986, both in control engineering. In 2005, he was awarded the higher doctorate degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK.

From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has

been with Electronics and Computer Science, the University of Southampton, UK, where he holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen's research interests include adaptive signal processing, wireless communications, modelling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimisation. He has published over 550 research papers.

Dr. Chen is a Fellow of IET. He is a Distinguished Adjunct Professor at the King Abdulaziz University, Jeddah, Saudi Arabia. He is an ISI highly cited researcher in the engineering category (March 2004). He was elected to a Fellow of the United Kingdom Royal Academy of Engineering in 2014.