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*— A 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 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 KICA (WKICA). In particular, in WKICA, GMM is first 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 online fault detection. A nonlinear contribution plots method is also developed based on the idea of a 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*— Contribution plots, fault detection, fault identification, Gaussian mixture model (GMM), kernel independent component analysis (KICA), process monitoring.**

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. A 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 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 with PCA, ICA can effectively utilize the higher order information in non-Gaussian process data and extracts mutually independent latent variables known as 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. Kano *et al.* [18] 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, Odiowei 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

Manuscript received March 9, 2015; revised November 16, 2015; accepted November 30, 2015. Date of publication December 17, 2015; date of current version December 22, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61273160 and Grant 61403418.

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

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2015.2505086

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] first introduced the kernel ICA (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 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 data set 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. However, 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. The main motivation of this paper 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 online 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 (TE) 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 liner feature space to obtain the feature data and whitening the feature data in the feature space using KPCA and 2) performing ICA on the whitened data in the KPCA-whitened space to extract the KICs features.

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

The covariance matrix $\boldsymbol{C} \in \boldsymbol{\mathcal{F}} \times \boldsymbol{\mathcal{F}}$ of the data $\bar{\boldsymbol{\Phi}}$ may be estimated by $\boldsymbol{C} = (1/n)\bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\Phi}}^T$. A straightforward idea would be to find the eigenvectors of $\boldsymbol{C}$ for obtaining the PCs in the feature space. However, the functional form of $\boldsymbol{\phi}(\cdot)$ is unknown, which makes it infeasible by the eigendecomposition of $\boldsymbol{C}$ directly. Hence, kernel trick is introduced to solve this problem. Define a Gram kernel matrix $\boldsymbol{K} \in \mathbb{R}^{n \times n}$ as

$$\boldsymbol{K} = \boldsymbol{\Phi}^T \boldsymbol{\Phi}. \tag{1}$$

The ith-row and jth-column elements of $\boldsymbol{K}$ can be written as

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

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

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

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 $\boldsymbol{\phi}(\cdot)$ explicitly.

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

$$\bar{\boldsymbol{K}} = \bar{\boldsymbol{\Phi}}^T\bar{\boldsymbol{\Phi}} = \boldsymbol{K} - \mathbf{1}_n\boldsymbol{K} - \boldsymbol{K}\mathbf{1}_n + \mathbf{1}_n\boldsymbol{K}\mathbf{1}_n \tag{4}$$

where $\mathbf{1}_n \in \mathbb{R}^{n\times n}$ denotes the matrix whose elements are all equal to $(1/n)$. Let $\lambda_i \in \mathbb{R}$, $1 \le i \le n$, be the eigenvalues of $\bar{\boldsymbol{K}}$ satisfying the condition $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$, and $\boldsymbol{\beta}_i \in \mathbb{R}^n$ for $1 \le i \le n$ be the corresponding eigenvectors of $\bar{\boldsymbol{K}}$. The estimate of the covariance matrix $\boldsymbol{C}$ can be expressed as

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

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

$$\boldsymbol{V} = \bar{\boldsymbol{\Phi}}\boldsymbol{H}_a\boldsymbol{\Lambda}_a^{-1/2} \tag{6}$$

where $\boldsymbol{\Lambda}_a = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_a\}$ and $\boldsymbol{H}_a = [\boldsymbol{\beta}_1\ \boldsymbol{\beta}_2\ \cdots\ \boldsymbol{\beta}_a] \in \mathbb{R}^{n\times a}$. Then, the kernel whitening matrix $\boldsymbol{Q}_F \in \mathbb{R}^a \times \boldsymbol{\mathcal{F}}$ can be obtained in the feature space $\boldsymbol{\mathcal{F}}$ as

$$\boldsymbol{Q}_F = \text{diag}\left\{\frac{\lambda_1}{n}, \frac{\lambda_2}{n}, \ldots, \frac{\lambda_a}{n}\right\}^{-1/2}\boldsymbol{V}^T = \sqrt{n}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T\bar{\boldsymbol{\Phi}}^T. \tag{7}$$

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

$$\boldsymbol{Z} = [\boldsymbol{z}_1\ \ \boldsymbol{z}_2\ \ \cdots\ \ \boldsymbol{z}_n] = \boldsymbol{Q}_F\bar{\boldsymbol{\Phi}} = \sqrt{n}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T\bar{\boldsymbol{K}} \tag{8}$$

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

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]:

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

where $\boldsymbol{z} \in \mathbb{R}^a$ denotes the vector of a whitened variables, $\boldsymbol{u}_i \in \mathbb{R}^a$ denotes the ith maximum non-Gaussian direction related to $\boldsymbol{z}$, υ denotes a Gaussian variable with zero mean and unit variance, and $E\{\cdot\}$ denotes the expectation operator, while $G(\cdot)$ is a nonquadratic function that can be chosen as $G(\boldsymbol{u}_i^T\boldsymbol{z}) = -\exp(-(\boldsymbol{u}_i^T\boldsymbol{z})^2/2)$. 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 $\boldsymbol{U} = [\boldsymbol{u}_1\ \boldsymbol{u}_2\ \cdots\ \boldsymbol{u}_a]^T$. It can easily be verified that the matrix $\boldsymbol{U}$ is an orthogonal matrix. Based on $\boldsymbol{U}$, the samples $\boldsymbol{S} \in \mathbb{R}^{a\times n}$ of the KICs can be extracted according to

$$\boldsymbol{S} = [\boldsymbol{s}_1\ \ \boldsymbol{s}_2\ \ \cdots\ \ \boldsymbol{s}_n] = \boldsymbol{U}\boldsymbol{Z}. \tag{10}$$

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

$$\boldsymbol{z}_t = \boldsymbol{Q}_F\bar{\boldsymbol{\phi}}(\boldsymbol{x}_t) = \sqrt{n}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T\bar{\boldsymbol{\Phi}}^T\bar{\boldsymbol{\phi}}(\boldsymbol{x}_t) = \sqrt{n}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T\bar{\boldsymbol{k}}_t \tag{11}$$

$$\boldsymbol{s}_t = \boldsymbol{U}\boldsymbol{z}_t \tag{12}$$

where $\bar{\boldsymbol{\phi}}(\boldsymbol{x}_t) \in \boldsymbol{\mathcal{F}}$ denotes the current zero-mean feature data in the feature space, and $\boldsymbol{z}_t \in \mathbb{R}^a$ denotes the current value of the whitened variable, while $\bar{\boldsymbol{k}}_t = \boldsymbol{k}_t - \boldsymbol{K}\mathbf{1}_1 - \mathbf{1}_n\boldsymbol{k}_t + \mathbf{1}_n\boldsymbol{K}\mathbf{1}_1$ with the vector $\mathbf{1}_1 \in \mathbb{R}^n$ whose elements are all equal to $1/n$, and $\boldsymbol{k}_t = [k(\boldsymbol{x}_1, \boldsymbol{x}_t)\ k(\boldsymbol{x}_2, \boldsymbol{x}_t)\ \cdots\ k(\boldsymbol{x}_n, \boldsymbol{x}_t)]^T$.

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

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

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

where $d < a$ is the number of the dominant KICs, $\boldsymbol{s}_{d,t} = [s_{1,t}\ s_{2,t}\ \cdots\ s_{d,t}]^T = \boldsymbol{U}_d\boldsymbol{z}_t$ with $\boldsymbol{s}_{d,t} \in \mathbb{R}^d$ contains the first d extracted KICs, called the dominant KICs at time t, and $\boldsymbol{U}_d \in \mathbb{R}^{d\times a}$ consists of the first d rows of $\boldsymbol{U}$, while $\boldsymbol{e}_t = \boldsymbol{z}_t - \boldsymbol{U}_d^T\boldsymbol{U}_d\boldsymbol{z}_t$ with $\boldsymbol{e}_t \in \mathbb{R}^a$ is the residual vector of the KICA model, $\boldsymbol{s}_{e,t} = [s_{d+1,t}\ s_{d+2,t}\ \cdots\ s_{a,t}]^T = \boldsymbol{U}_e\boldsymbol{z}_t$ with $\boldsymbol{s}_{e,t} \in \mathbb{R}^{a-d}$ contains the last $a-d$ extracted KICs, called the excluded KICs at time t, and $\boldsymbol{U}_e \in \mathbb{R}^{(a-d)\times a}$ consists of the last $a-d$ rows of $\boldsymbol{U}$. 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 nonsystematic part change in the residual of the KICA model [6], [13]. We determine the hyperdimension 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 that 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 online monitoring, an effective means for improving fault detection performance is to emphasize those KICs that 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 that 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:

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

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

where WI_t^2 and WQ_t are the improved versions of the monitoring statistics I_t^2 and Q_t given in (13) and (14), respectively, while $w_{i,t}$ is the weight that is determined according to the importance degree of the ith 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. In particular, 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 data sets 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 tradeoff 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, Santamaría *et al.* [24] specifically designed an improved expectation–maximization (EM) algorithm.

Thus, for the ith 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:

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

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

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

The parameters ξ_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 $\boldsymbol{S} = [\boldsymbol{s}_1\ \boldsymbol{s}_2\ \cdots\ \boldsymbol{s}_n]$ with $\boldsymbol{s}_t = [s_{1,t}\ s_{2,t}\ \cdots\ s_{a,t}]^T$ for $1 \le t \le n$. Then, the parameters of the two-Gaussian mixture are estimated iteratively according to

$$\xi_i\big|^{l+1} = \gamma\, \xi_i\big|^l + \frac{1-\gamma}{n} \sum_{t=1}^{n} \Theta_{i,t}\Big|^l \tag{19}$$

with

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

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

$$\sigma_{i,2}^2\big|^{l+1} = \frac{1-\xi_i\big|^{l+1}\sigma_{i,1}^2\big|^{l+1}}{1-\xi_i\big|^{l+1}} \tag{22}$$

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

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

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

where δ 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 lookup table of the standard Gaussian cdf can be stored, and the task of online computing $f_i(s_{t,i})$ becomes the one that simply retrieves the four points of the standard Gaussian cdf from the lookup 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 $f_i(s_{i,t})$, $f_i(s_{i,t-1}), \ldots, f_i(s_{i,t-q+1})$ over q samples can be calculated according to

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

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:

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

where $0 < \eta < 0.5$ is a prespecified value, and $\bar{f}_{i,\lim}$ is a probability threshold for determining whether the KIC s_i is important for reflecting the useful process information at the time t. If $\bar{f}_{i,t} > \bar{f}_{i,\lim}$, the sample $s_{i,t}$ is considered to be within the normal operating region and contains no fault information. Then, a smaller weight value η is given to the sample $s_{i,t}$ to suppress the irrelevant information provided by $s_{i,t}$. On the other hand, the condition $\bar{f}_{i,t} \le \bar{f}_{i,\lim}$ 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 η 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 η. In particular, it is suggested to set the value of η 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 η 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 $\bar{f}_{i,\lim}$ 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 ith KIC s_i, the corresponding probabilities $\{\bar{f}_{i,t}\}_{t=q}^{n}$ can be used to help choosing an appropriate value for $\bar{f}_{i,\lim}$. In particular, for the given α confidence limit, $(n-q+1)(1-\alpha)$ is rounded toward the nearest integer, denoted by r, and then, the rth lowest value of $\{\bar{f}_{i,t}\}_{t=q}^{n}$ is chosen as the probability threshold $\bar{f}_{i,\lim}$.

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 α confidence limit $WI_{\lim,\alpha}^2$ for the monitoring statistic WI_t^2 and the α confidence limit $WQ_{\lim,\alpha}$ 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 following offline modeling stage.

1) Collect the data from the process under normal operating conditions, and divide the process data into the training data set and the validating data set.
2) Based on the training data, construct the KICA model to obtain the training samples of the KICs.
3) Fit the pdf of each KIC by using the GMM (17) and estimate the GMM parameters based on the training samples of each KIC using (19)–(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 η 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 α 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 prespecified reasonable confidence range, the value of η 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 proposed fault detection strategy using the GMM-based WKICA includes the following online 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

TABLE I
COMPARISON OF THE ONLINE 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	$da-1$	$(a-d)a-1$	$da-1$	$(a-d)a-1$

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 a nonlinear structure, while the WICA is only suitable for dealing with non-Gaussian data with a 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 the WICA adopt KDE to estimate pdf. KDE uses all the n training samples to estimate the pdf of a new sample online. In general, the training data size n should be sufficiently large in order to guarantee the accuracy of density estimation. Thus, the complexity of online calculating the corresponding probability weight values based on the KDE approach may become unacceptably high. Various sparse KDE methods [32]–[34] may be adopted instead to provide sparse estimates of pdf in order to alleviate the high online computation burden. From this viewpoint, our GMM-based pdf estimation, which uses the two-Gaussian mixture to estimate pdf, minimizes the online computational complexity in a real-time fault detection application, while having a desired capability of modeling non-Gaussian pdf.

1) Online Fault Detection Complexity: The online 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 online 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 online based on (23). Since the values of standard Gaussian cdf are calculated offline and stored in the memory space, the online computing of the probabilities $f_i(s_{i,t})$ for $1 \le i \le a$ 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 a few existing works discussed the methods of identifying faults. In particular, there is hardly any fault identification research for the KICA-based monitoring methods owing to the usage of implicit nonlinear transformation. Based on the idea of a sensitivity analysis [25], we develop a novel nonlinear contribution plots method for fault identification. In particular, define

$$\boldsymbol{C}_{WI_t^2} = \frac{\partial WI_t^2}{\partial \boldsymbol{x}_t} \circ \boldsymbol{x}_t \tag{26}$$

$$\boldsymbol{C}_{WQ_t} = \frac{\partial WQ_t}{\partial \boldsymbol{x}_t} \circ \boldsymbol{x}_t \tag{27}$$

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

First, 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. Second, 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 online operation. Third, 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 $\{\boldsymbol{C}_{WI_t^2}\}_{t=1}^n$ and $\{\boldsymbol{C}_{WQ_t}\}_{t=1}^n$ 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 $\{\boldsymbol{C}_{WI_t^2}\}_{t=t_{\min}}^{t_{\min}+L-1}$ and $\{\boldsymbol{C}_{WQ_t}\}_{t=t_{\min}}^{t_{\min}+L-1}$, and scale them with the means and variances obtained in Step 1). Denote the absolute values of the normalized contribution series by $\{|\bar{\boldsymbol{C}}_{WI_t^2}|\}_{t=t_{\min}}^{t_{\min}+L-1}$ and $\{|\bar{\boldsymbol{C}}_{WQ_t}|\}_{t=t_{\min}}^{t_{\min}+L-1}$.
4) The means of $\{|\bar{\boldsymbol{C}}_{WI_t^2}|\}_{t=t_{\min}}^{t_{\min}+L-1}$ provide the contributions of the process variables to the monitoring statistic WI_t^2, while the means of $\{|\bar{\boldsymbol{C}}_{WQ_t}|\}_{t=t_{\min}}^{t_{\min}+L-1}$ 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. Four-Variable System

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

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

where $\boldsymbol{x} = [x_1\ x_2\ x_3\ x_4]^T \in \mathbb{R}^4$ contained the four nonlinear and non-Gaussian output variables measured for fault detection

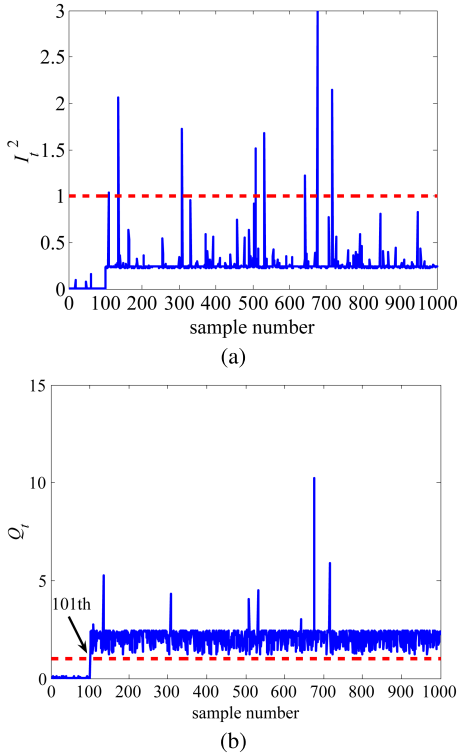


Fig. 1. (a) I_t^2 monitoring chart and (b) Q_t monitoring chart of the KICA method for the four-variable system under fault pattern 1. Dashed line: 99% confidence limit for the corresponding monitoring statistic.

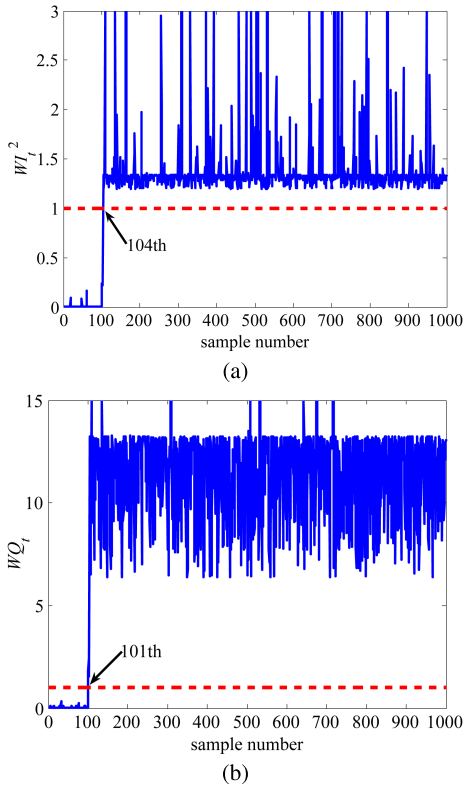


Fig. 2. (a) WI_t^2 monitoring chart and (b) WQ_t monitoring chart of the WKICA method for the four-variable system under fault pattern 1. Dashed line: 99% confidence limit for the corresponding monitoring statistic.

and fault identification, while $b = 0.3b_0$ and b_0 was the non-Gaussian input variable whose pdf was given by

$$p(b_0) = \xi g_0\left(b_0; \sigma_1^2\right) + (1-\xi) g_0\left(b_0; \sigma_2^2\right) \tag{29}$$

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.

1) *Fault 1:* A step change in the process variable x_4 with the step changing value of -0.15.
2) *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 statistic values exceed the confidence limit to all the fault samples [7], [10], [14], [16], [19]. The hyperdimension a, the number of dominant KICs d, and the kernel width c were found according to the empirical rule suggested in [14]. In particular, a was determined as the number of the eigenvalues of the matrix $\bar{\boldsymbol{K}}$, which satisfies the condition

$$\lambda_i \Big/ \sum_{j=1}^{n} \lambda_j > 0.0001, \quad i \in \{1, 2, \ldots, n\}. \tag{30}$$

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 cutoff 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 η in (25) was set to 0.3, and the corresponding false alarm rates calculated based on the validation data set of the normal operation[1] 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 η.

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 dashed 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

[1]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 η to all the extracted KICs, i.e., it will use an equal weighting strategy similar to the KICA method.

TABLE II

FALSE ALARM RATES (%) ASSOCIATED WITH THE TWO MONITORING STATISTICS OF EACH METHOD FOR THE VALIDATION DATA SET 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

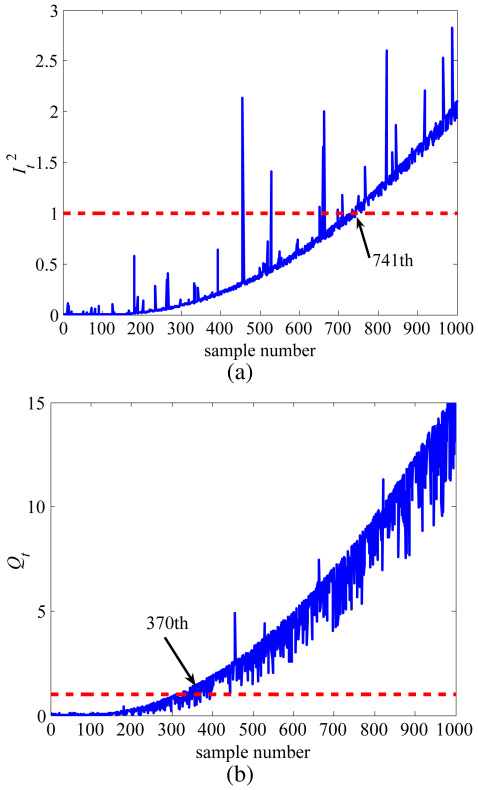


Fig. 3. (a) I_t^2 monitoring chart and (b) Q_t monitoring chart of the KICA method for the four-variable system under fault pattern 2. Dashed line: 99% confidence limit for the corresponding monitoring statistic.

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 a 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 fault pattern 2 are shown in Figs. 3 and 4, respectively. Clearly, the WKICA method has a superior fault detection performance over the KICA method for this fault pattern. In particular, the two monitoring charts of the WKICA method are able to detect the occurrence of the fault much earlier than the two corresponding monitoring charts of the KICA method. Moreover, the WI_t^2 monitoring chart attains

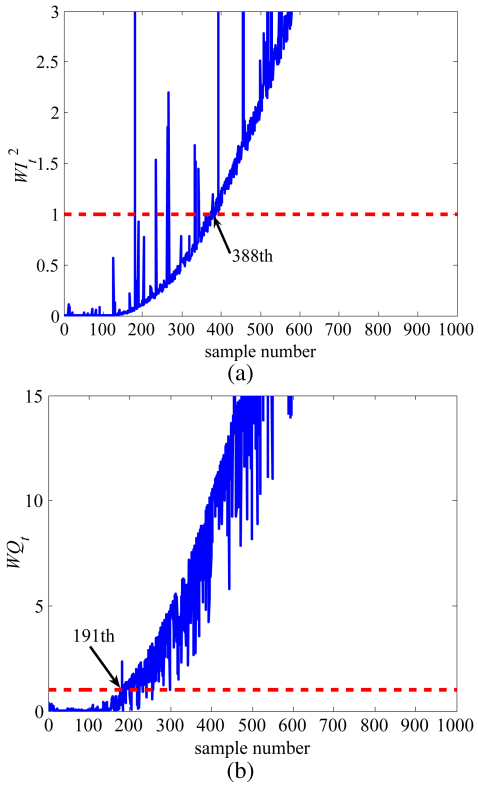


Fig. 4. (a) WI_t^2 monitoring chart and (b) WQ_t monitoring chart of the WKICA method for the four-variable system under fault pattern 2. Dashed line: 99% confidence limit for the corresponding monitoring statistic.

a 70% fault detection rate, compared with the 30% fault detection rate of the I_t^2 monitoring chart, while the WQ_t monitoring chart achieves a 89.67% fault detection rate, in comparison with the 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 the blue solid curve denotes the estimated probabilities of the KICs' samples from the training data, and the magenta dotted-dashed curve represents the estimated probabilities of the KICs' samples from fault pattern 2 data, while the red dashed 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 data set 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 η according to (25). By contrast, the magenta dotted-dashed 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. In particular, 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

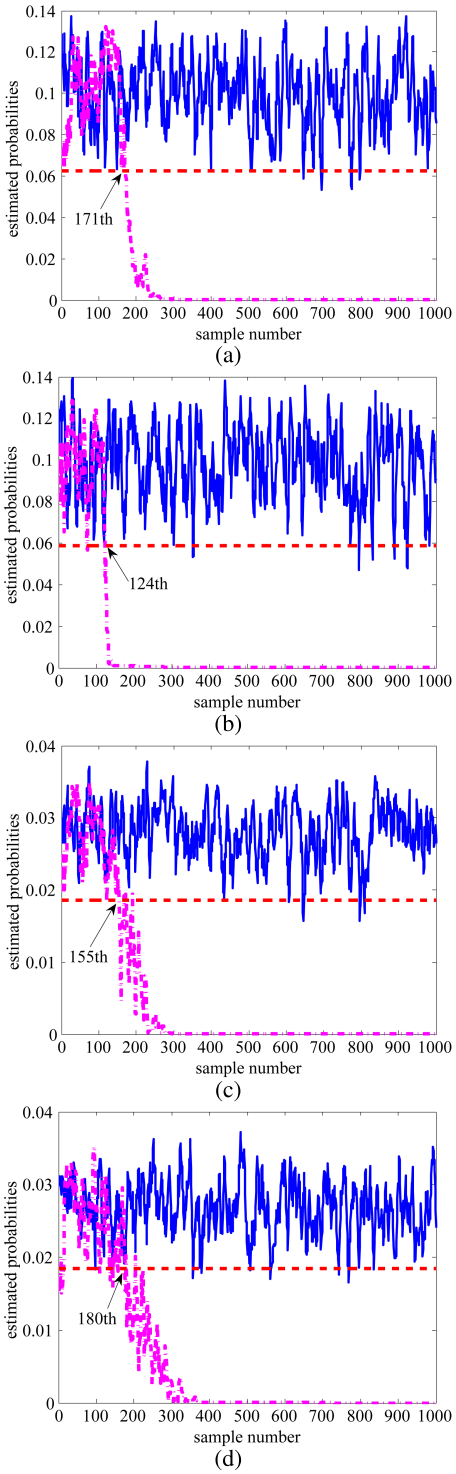


Fig. 5. Estimated $\bar{f}_{i,t}$ for the samples of s_i. (a) $i = 1$, (b) $i = 2$, (c) $i = 3$, and (d) $i = 4$ for the training data (blue solid curve) and the data of fault pattern 2 (magenta dotted-dashed curve). Dashed line: probability threshold for the corresponding KIC.

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

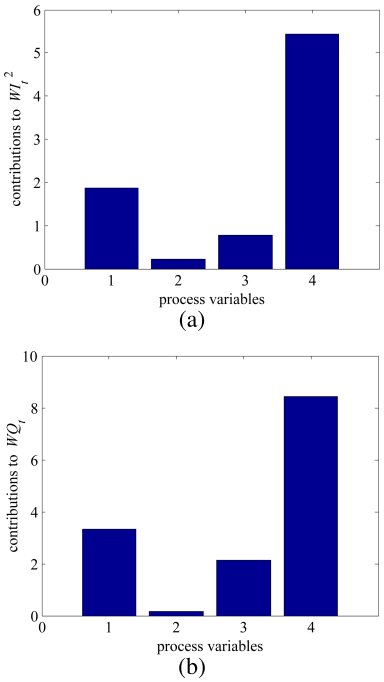


Fig. 6. Fault identification results using our nonlinear contribution plots method for the four-variable system under fault pattern 1. (a) Process variables' contributions to WI_t^2. (b) Process variables' contributions to WQ_t.

fault identification is an unsolved open problem, especially for the nonlinear ICA-based monitoring methods with a kernel technique. Here, we demonstrate the potential of our proposed nonlinear contribution plots method in fault identification. For 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 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. Tennessee Eastman Industrial Process

The 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 shown 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

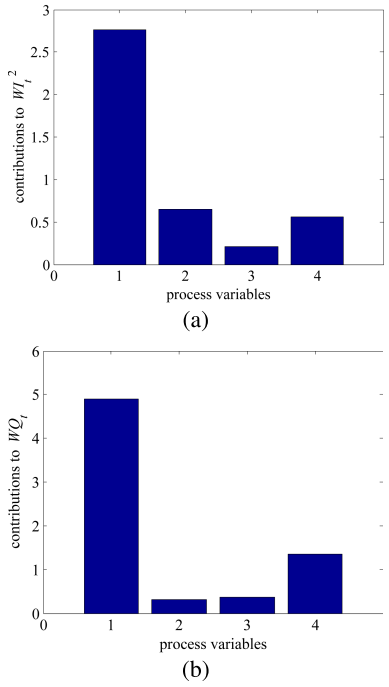


Fig. 7. Fault identification results using our nonlinear contribution plots method for the four-variable system under fault pattern 2. (a) Process variables' contributions to WI_t^2. (b) Process variables' contributions to WQ_t.

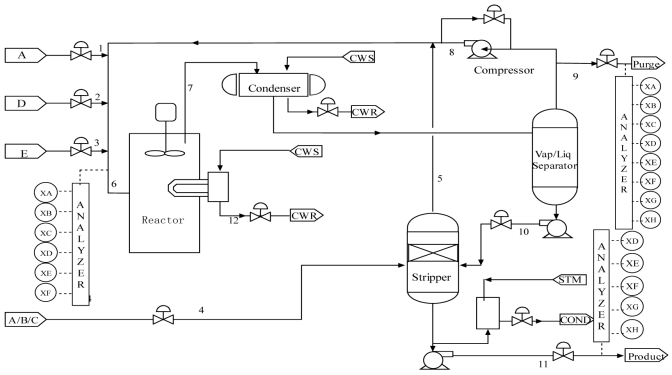


Fig. 8. Flowchart of the TE industrial process.

monitoring this process, including 11 manipulated variables, 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 preprogrammed process faults, which are listed in Table IV. Each fault pattern contains 960 samples with a sampling interval of 3 min, and the corresponding fault is introduced at the 160th sample. The TE simulator also produces two normal operating data sets with 480 samples and 960 samples, respectively.

The normal operating data set with 960 samples was used for training, while the other normal operating data set with 480 samples was adopted as the validating set for testing the false alarm rate of the built model. The hyperdimension 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,

TABLE III
MEASURED PROCESS VARIABLES OF THE TE 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 TE 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

yielding $c = 6000$, $a = 54$, and $d = 42$. Thus, the first 42 KICs s_i, $1 \le i \le 42$, were used in constructing the I_t^2 and WI_t^2 monitoring charts, while the last 12 KICs s_i, $43 \le i \le 52$, 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

TABLE V

FALSE ALARM RATES (%) ASSOCIATED WITH THE TWO MONITORING STATISTICS OF EACH METHOD FOR THE TE 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

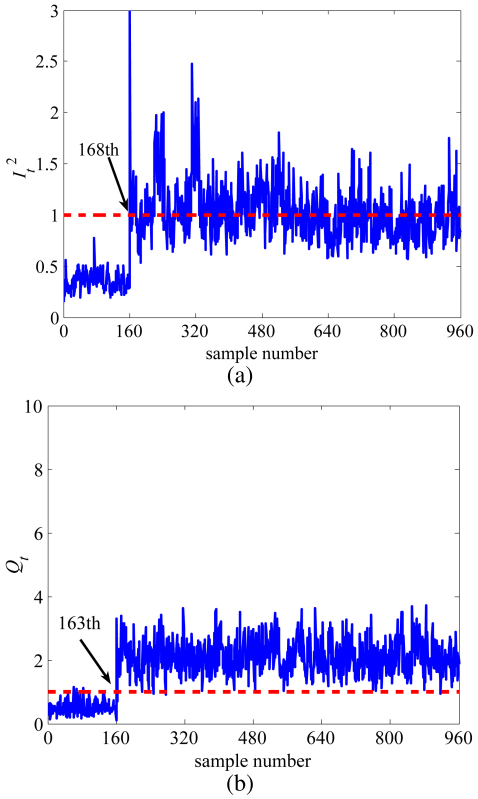


Fig. 9. (a) I_t^2 monitoring chart and (b) Q_t monitoring chart of the KICA method for the TE process under fault pattern 4. Dashed line: 99% confidence limit for the corresponding monitoring statistic.

monitoring statistics, the parameter η was set to 0.3, and the estimated probabilities were averaged 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 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) with 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 not 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 Figs. 9(b) and 10(b) that, although both the Q_t and WQ_t monitoring charts can detect the fault at the 163rd sample

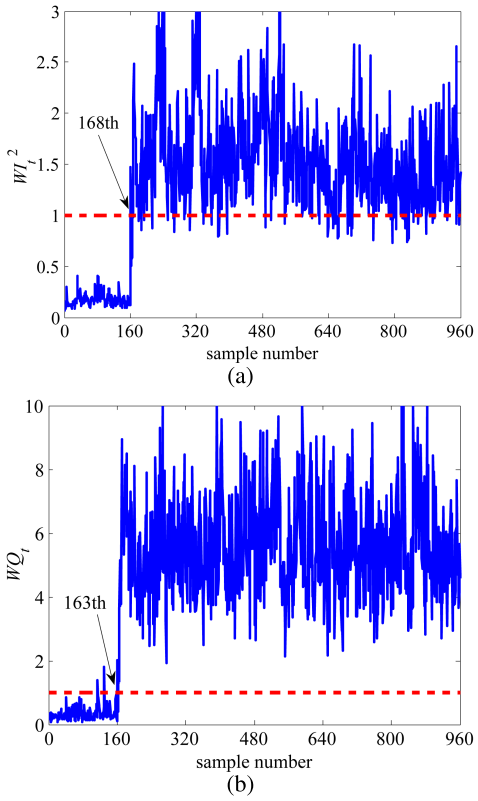


Fig. 10. (a) WI_t^2 monitoring chart and (b) WQ_t monitoring chart of the WKICA method for the TE process under fault pattern 4. Dashed line: 99% confidence limit for the corresponding monitoring statistic.

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 a fault detection time and a fault detection rate. 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 data sets [35]. Therefore, we excluded these three fault patterns in our investigation. From Table VI, it is seen that the both methods achieve similar fault detection times for most of the 18 fault patterns, but there are three fault patterns for which the monitoring charts of our WKICA method detect the occurrence of fault much earlier than the monitoring charts of the KICA method. A fault detection time does not tell how reliable the detection is, and for measuring the reliability of fault detection, we need to turn to a 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, both the 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

TABLE VI

FAULT DETECTION TIMES (SAMPLE NUMBER) ACHIEVED BY THE KICA-BASED AND WKICA-BASED METHODS FOR 18 FAULT PATTERNS OF THE TE 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	**256**	170	**171**	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	**failed**	**195**	**333**	**171**
20	245	241	242	241
21	**675**	**711**	**662**	**665**

TABLE VII

FAULT DETECTION RATES (%) ACHIEVED BY THE KICA-BASED AND WKICA-BASED METHODS FOR 18 FAULT PATTERNS OF THE TE 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	**49.13**	99.38	**91.62**	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	**72.63**	**76.38**	**79.12**	**84.25**
11	**42.75**	**73.25**	**50.88**	**82.00**
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	**71.75**	**84.00**	**78.50**	**90.38**
17	**86.38**	96.13	**91.62**	97.25
18	89.63	90.38	89.62	90.50
19	**41.25**	**77.38**	**54.25**	**89.38**
20	**47.88**	**57.50**	**54.75**	**67.62**
21	**36.38**	**34.38**	**40.00**	**43.62**

detecting complex faults, including 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 fault patterns 6 and 11. For 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 fault pattern 6, where it can be clear seen that the 1st and 44th process variables have the highest contributions to this fault. Fault pattern 6 is a step-type fault, which is caused by the component A feed loss in the stream 1, as shown 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,

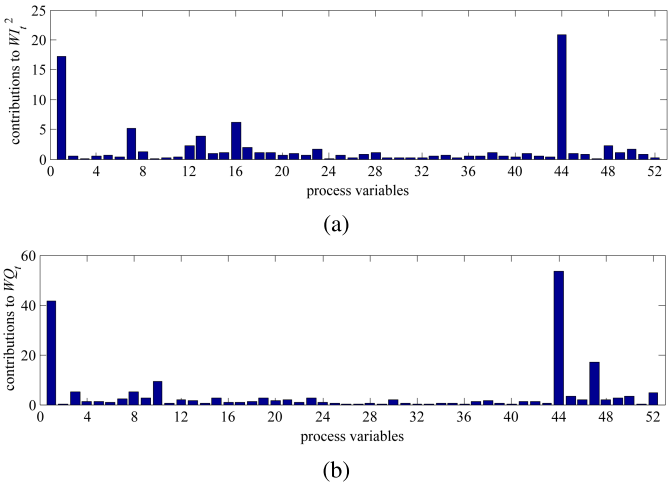


Fig. 11. Fault identification results using our nonlinear contribution plots method for the TE process under fault pattern 6. (a) Process variables' contributions to WI_t^2. (b) Process variables' contributions to WQ_t.

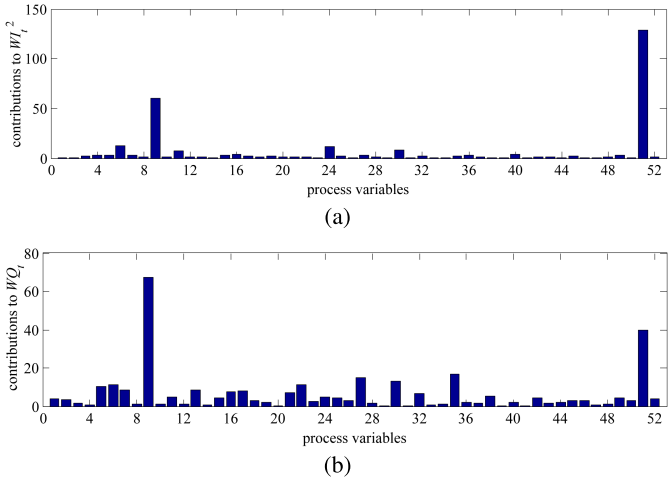


Fig. 12. Fault identification results using our nonlinear contribution plots method for the TE process under fault pattern 11. (a) Process variables' contributions to WI_t^2. (b) Process variables' contributions to WQ_t.

and therefore, they provide the direct and effective guidance for locating the fault source.

For 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 shown 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, fault pattern 11 is caused by the random variation of the reactor cooling water inlet temperature, according to the fault description given in Table III. However, the reactor cooling water inlet temperature is not contained in the set of the TE process's monitored 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 provide the plant operator who has the underlying plant knowledge with the valuable information to locate the root of the fault.

V. Conclusion

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 TE benchmark process have demonstrated that the proposed GMM-based WKICA method outperforms the existing state-of-the-art KICA method, in terms of a false detection time and a 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 is two open problems. Further works are warranted to provide theoretical direction and practical implementation.

Appendix

Noting $\boldsymbol{s}_{d,t} = \boldsymbol{U}_d\sqrt{n}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T\bar{\boldsymbol{k}}_t$, the monitoring statistic WI_t^2 can be expressed as

$$WI_t^2 = \bar{\boldsymbol{k}}_t^T\boldsymbol{M}_1\bar{\boldsymbol{k}}_t \tag{31}$$

where

$$\boldsymbol{M}_1 = n\boldsymbol{H}_a\boldsymbol{\Lambda}_a^{-1}\boldsymbol{U}_d^T\mathrm{diag}\{w_{1,t}^2,\ldots,w_{d,t}^2\}\boldsymbol{U}_d\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T. \tag{32}$$

Similarly, since $\boldsymbol{s}_{e,t} = \boldsymbol{U}_e\sqrt{n}\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T\bar{\boldsymbol{k}}_t$, we have

$$WQ_t = \bar{\boldsymbol{k}}_t^T\boldsymbol{M}_2\bar{\boldsymbol{k}}_t \tag{33}$$

where

$$\boldsymbol{M}_2 = n\boldsymbol{H}_a\boldsymbol{\Lambda}_a^{-1}\boldsymbol{U}_e^T\mathrm{diag}\{w_{d+1,t}^2,\ldots,w_{a,t}^2\}\boldsymbol{U}_e\boldsymbol{\Lambda}_a^{-1}\boldsymbol{H}_a^T. \tag{34}$$

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

$$\begin{aligned}\boldsymbol{C}_{WI_t^2} &= \left(\left(\left[\frac{\partial WI_t^2}{\partial\bar{\boldsymbol{k}}_t}\cdots\frac{\partial WI_t^2}{\partial\bar{\boldsymbol{k}}_t}\right]^T\right.\right.\\ &\qquad\left.\left.\circ\left[\frac{\partial\bar{\boldsymbol{k}}_t}{\partial x_{1,t}}\cdots\frac{\partial\bar{\boldsymbol{k}}_t}{\partial x_{m,t}}\right]^T\right)\times\boldsymbol{one}_n\right)\circ\boldsymbol{x}_t\\ &= \left(\left([2\boldsymbol{M}_1\bar{\boldsymbol{k}}_t\cdots 2\boldsymbol{M}_1\bar{\boldsymbol{k}}_t]^T\right.\right.\\ &\qquad\left.\left.\circ\left[\frac{\partial\bar{\boldsymbol{k}}_t}{\partial x_{1,t}}\cdots\frac{\partial\bar{\boldsymbol{k}}_t}{\partial x_{m,t}}\right]^T\right)\times\boldsymbol{one}_n\right)\circ\boldsymbol{x}_t\end{aligned} \tag{35}$$

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

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

$$\begin{aligned}\boldsymbol{C}_{WQ_t} &= \left(\left([2\boldsymbol{M}_2\bar{\boldsymbol{k}}_t\cdots 2\boldsymbol{M}_2\bar{\boldsymbol{k}}_t]^T\right.\right.\\ &\qquad\left.\left.\circ\left[\frac{\partial\bar{\boldsymbol{k}}_t}{\partial x_{1,t}}\cdots\frac{\partial\bar{\boldsymbol{k}}_t}{\partial x_{m,t}}\right]^T\right)\times\boldsymbol{one}_n\right)\circ\boldsymbol{x}_t.\end{aligned} \tag{36}$$

The derivatives $(\partial\bar{\boldsymbol{k}}_t/\partial x_{j,t})$, $1\le j\le m$, are given by

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

in which

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

References

[1] S. J. Qin and Y. Y. Zheng, "Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures," *AIChE J.*, vol. 59, no. 2, pp. 496–504, Feb. 2013.

[2] X. Tian, L. Cai, and S. Chen, "Noise-resistant joint diagonalization independent component analysis based process fault detection," *Neurocomputing*, vol. 149, pp. 652–666, Feb. 2015.

[3] J. Zeng, L. Xie, U. Kruger, and C. Gao, "Regression-based analysis of multivariate non-Gaussian datasets for diagnosing abnormal situations in chemical processes," *AIChE J.*, 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," *Chem. Eng. Sci.*, 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 Intell. Lab. Syst.*, 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 Intell. Lab. Syst.*, 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 modelling approach," *Comput. Chem. Eng.*, 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 Intell. Lab. Syst.*, 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 Process.*, 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 Intell. Lab. Syst.*, 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," *Appl. Soft Comput.*, vol. 14, 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.

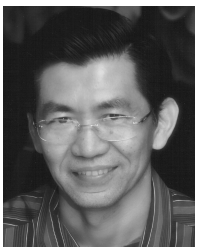
[13] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AIChE J.*, vol. 52, no. 10, pp. 3501–3514, Oct. 2006.
[14] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection of non-linear processes using kernel independent component analysis," *Can. J. Chem. Eng.*, 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," *Chem. Eng. Sci.*, 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," *Ind. Eng. Chem. Res.*, 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 J.*, 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 Intell. Lab. Syst.*, 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," *Ind. Eng. Chem. Res.*, 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 Recognit.*, 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 Eng. Pract.*, 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. Geosci. Remote Sens.*, 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," *Comput. Chem. Eng.*, vol. 30, nos. 10–12, pp. 1553–1559, Sep. 2006.
[26] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, 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 Anal. Mach. Intell.*, 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," *Comput. Chem. Eng.*, 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," *Comput. Chem. Eng.*, 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 Process.*, 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 Intell. Lab. Syst.*, vol. 127, pp. 121–131, Aug. 2013.
[32] M. Girolami and C. He, "Probability density estimation from optimally condensed data samples," *IEEE Trans. Pattern Anal. Mach. Intell.*, 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 Netw.*, 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 modeling," *IEEE Trans. Evol. Comput.*, 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*. London, U.K.: Springer-Verlag, 2001.

Lianfang Cai received the B.Eng. and Ph.D. degrees from the China University of Petroleum, Beijing, China, in 2009 and 2014, respectively.

Since 2014, he has been with the College of Information and Control Engineering, China University of Petroleum. Currently he is also a Post-Doctoral Researcher with Imperial College London, U.K. His current research interests include data-driven fault detection and diagnosis, multivariate statistical analysis, and blind signal processing.

Xuemin Tian received the B.Eng. degree from the Huadong Petroleum Institute, Dongying, China, in 1982, and the M.S. degree from the China University of Petroleum, Beijing, China, in 1994.

He served as a Visiting Professor of Process Control with the University of California at Santa Barbara, Santa Barbara, CA, USA, from 2001 to 2002. He is currently a Professor of Process Control with the China University of Petroleum. His current research interests include modeling, advanced process control and optimization for petrol-chemical processes as well as fault detection and diagnosis, and process monitoring.



Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from the East China Petroleum Institute, Beijing, China, in 1982, the Ph.D. degree in control engineering from City University London, London, U.K., in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

He held research and academic appointments with the University of Sheffield, Sheffield, U.K., the University of Edinburgh, Edinburgh, U.K., and the University of Portsmouth, Portsmouth, U.K., from 1986 to 1999. He was an ISI Highly Cited Researcher in the engineering category in 2004. Since 1999, he has been with the Electronics and Computer Science Department, University of Southampton, where he is currently a Professor of Intelligent Systems and Signal Processing. He is a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He has authored over 550 research papers. His current research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, and evolutionary computation methods and optimization.

Dr. Chen is a fellow of the Institution of Engineering and Technology. He was elected as a fellow of the Royal Academy of Engineering, U.K., in 2014.