Control Engineering Practice 18 (2010) 242–253

Contents lists available at ScienceDirect

Control Engineering Practice

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

A novel process monitoring approach with dynamic independent component analysis

Chun-Chin Hsu [a,*], Mu-Chen Chen [b], Long-Sheng Chen [c]

[a] Department of Industrial Engineering and Management, Chaoyang University of Technology, 168 Jifong E. Rd., Wufong Township Taichung County 41349, Taiwan
[b] Institute of Traffic and Transportation, National Chiao Tung University, 114 Chung Hsiao W. Rd., Sec. 1, Taipei 10012, Taiwan
[c] Department of Information Management, Chaoyang University of Technology, 168 Jifong E. Rd., Wufong Township Taichung County 41349, Taiwan

ARTICLE INFO

Article history:
Received 7 November 2008
Accepted 1 November 2009

Keywords:
PCA
ICA
Tennessee Eastman process
TPC
Adjusted outlyingness

ABSTRACT

A novel process monitoring scheme is proposed to compensate for shortcomings in the conventional independent component analysis (ICA) based monitoring method. The primary idea is first to augment the observed data matrix in order to take the process dynamic into consideration. An outlier rejection rule is then proposed to screen out outliers, in order to better describe the majority of the data. Finally, a rectangular measure is used as a monitoring statistic. The proposed approach is investigated via three cases: a simulation example, the Tennessee Eastman process and a real industrial case. Results indicate that the proposed method is more efficient as compared to alternate methods.

1. Introduction

Principal component analysis (PCA) has been a widely used technique for monitoring multivariate processes. However, PCA assumes that latent variables follow a Gaussian distribution. Martin and Morris (1996) reported that PCA extracted variables rarely conform to a multivariate Gaussian distribution in real industrial processes such as chemical and biological plants. More recently, independent component analysis (ICA) has been developed to deal with non-Gaussian process monitoring. Kano, Tanaka, Hasebe, Hashimoto, and Ohno (2003) developed ICA-based statistical process control (SPC), and showed its superiority over the PCA-based SPC. However, their proposed method applies control charts to individually monitor extracted ICA components, and it may produce false alarms. Thus, Lee, Yoo, and Lee (2004a) developed ICA-based monitoring statistics and variable contribution plots for process monitoring and diagnosis, respectively. Further, Lee, Qin, and Lee (2006) proposed a modified ICA algorithm to sort the proper order of ICA components and determine how many components should be extracted. Ge and Song (2007) proposed a new monitoring scheme by integrating ICA and PCA. Then Ge and Song (2008) developed an adaptive local model approach for monitoring nonlinear multiple mode processes. Readers can refer to this work on ICA-based monitoring methods in Yoo, Lee, Vanrolleghem, and Lee (2004), Lee, Yoo, and Lee (2004b), Xia and Howell (2004), González and Sánchez (2007), Lee, Qin, and Lee (2007), Lu, Wu, Keng, and Chiu (2008), and Zhu, Hong, Wong, and Wang (2008) .

The above-mentioned studies have demonstrated ICA to be an efficient tool for monitoring non-Gaussian processes. Yet there are still some disadvantages with the traditional ICA algorithm. ICA assumes observations at one time to be independent over time. This assumption is invalid because of dynamic process characteristics. Although Lee et al. (2004b) proposed a dynamic ICA (DICA) algorithm in order to deal with the dynamic non-Gaussian multivariate process, the DICA algorithm still has some limitations. First, the training dataset is assumed to be “clean”, which means there is no contamination (outliers) in the training dataset. The effect of outliers may lead to an incorrect conclusion, such as a wrong estimation of parameters. Further, DICA applied Mahalanobis-type distance as the monitoring statistic, in which all variables are assumed to be obtained from an elliptical distribution, particularly the multivariate Gaussian (Hubert & Van der Veeken, 2008). Nevertheless, the extracted ICA components usually exhibit skewed distributions. In order to illustrate the influence, consider two independent uniform source signals with range $[-1, 1]$; Fig. 1(a) shows its scatter plot. By mixing the source signals with matrix A, the scatter plot of observed signals is shown in Fig. 1(b), and it exhibits a high correlation between variables. Fig. 1(c) shows the scatter plot of ICA extracted components. Clearly, ICA can reconstruct the source signals and make variables be independent. For process monitoring purposes, two types of control limits (i.e., rectangular and elliptical) are also drawn with a dotted line in Fig. 1(c). Obviously, the rectangular

* Corresponding author. Tel.: +886 4 23323000.
E-mail address: cchsu@cyut.edu.tw (C.-C. Hsu).

0967-0661/$ - see front matter
doi:10.1016/j.conengprac.2009.11.002

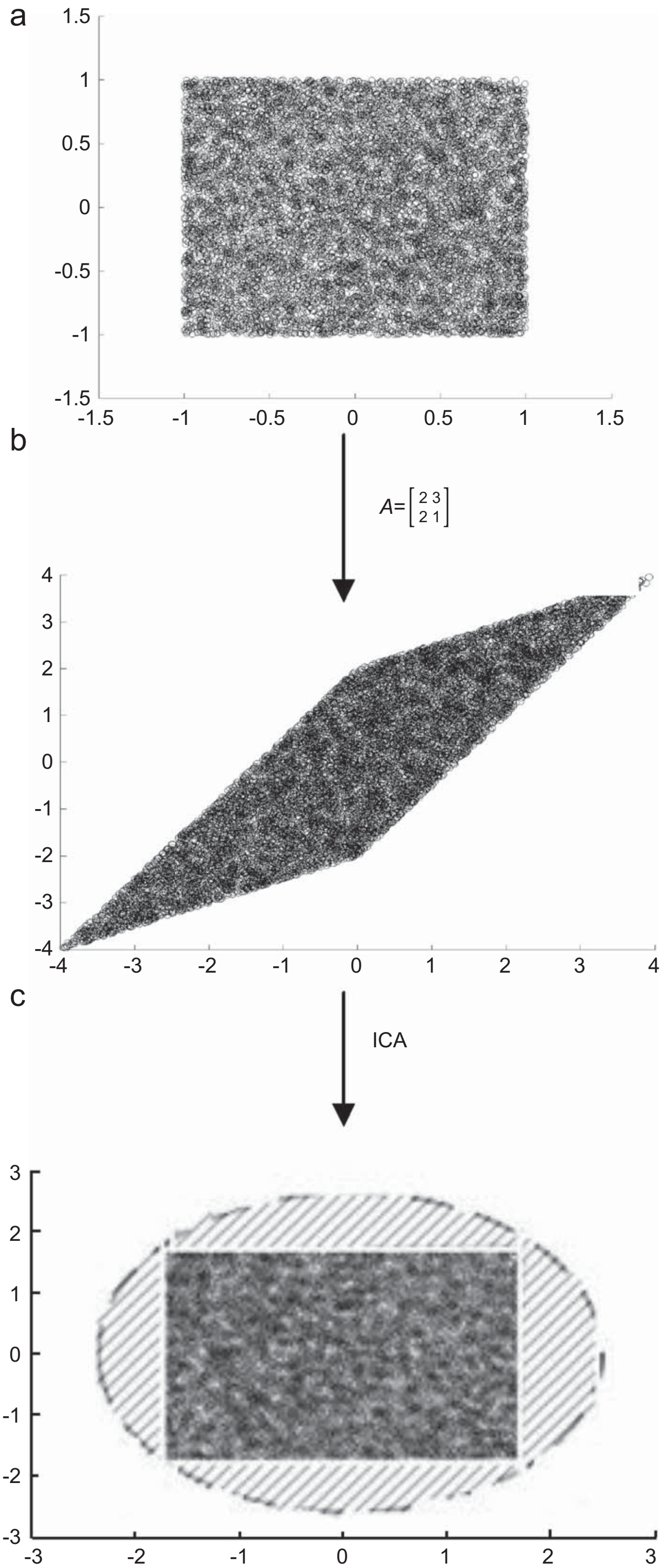

Fig. 1. (a) Source signal; (b) observed signal; and (c) ICA extracted signal.

type control limit fits better than the elliptical one. This implies that when an elliptical control limit is used for monitoring ICA components, the type II error (oblique lines in Fig. 1(c)) is increased and hence the fault detection rate is decreased. Note that the rectangular control limit only fits well for the assumption of variables to be skew-distributed. It implies that the elliptical control limit is still suitable for monitoring Gaussian processes. Thus, the constraint of this study assumes process variables to be non-Gaussian distributed.

This study proposes a new process monitoring scheme for ICA. The observed data matrix is first augmented by adding time-lagged variables for each measurement so as to take the process dynamic into consideration. After that, ICA is used for dimension reduction and then a rejection rule is proposed for excluding outliers. Next, the extracted ICA components are combined into a rectangular type monitoring measure. Finally, the kernel density estimation (KDE) method is utilized to determine the control limit. The proposed monitoring method will be investigated by using a simulated dynamical process with five variables and the Tennessee Eastman (TE) process. Additionally, it is applied to a real industrial case of a thermal power plant in Taiwan. To demonstrate the efficiency of this proposed method, several traditional monitoring methods are applied as benchmarks.

The remainder of this article is as follows. In the next section, the theory of the ICA algorithm is first introduced. After that, the ICA-based monitoring method is also introduced. The proposed method is then demonstrated in Section 3. Section 4 implements the proposed method and illustrates the comparisons with other alternatives. Finally, conclusions are drawn in Section 5.

2. ICA-based process monitoring

This section introduces the application of the ICA technique for non-Gaussian multivariate process monitoring. The theory of ICA is first reviewed, and an ICA-based process monitoring method which was proposed by Lee et al. (2004a) is then introduced.

2.1. ICA algorithm

In the ICA algorithm, the d observed variables $x_1, x_2, \ldots, x_d$ can be expressed in linear combination with m $(\leq d)$ unknown independent components $s_1, s_2, \ldots, s_m$. The relationship between them is given as

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where $\mathbf{X} \in \mathbf{R}^{d \times n}$ is the data matrix (unlike PCA, ICA employs the transposed data matrix), $\mathbf{S}$ is the independent component matrix, and $\mathbf{A}$ is the unknown mixing matrix.

The objective of ICA is to find a de-mixing matrix $\mathbf{W}$ such that the reconstructed signal $\hat{\mathbf{S}} = \mathbf{WX}$ becomes as independent as possible. The initial step in ICA is whitening. Assume the whitened signal can be expressed as $\mathbf{z} = \mathbf{QX}$ where $\mathbf{Q}$ denotes the whitening matrix. The $\mathbf{Q}$ can be obtained by calculating $\mathbf{Q} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of the data covariance matrix (i.e. $E(\mathbf{XX}^T)$) and U is a matrix with the corresponding eigenvectors as its columns. Thus, the whitened signal can be further expressed as

$$\mathbf{z} = \mathbf{QX} = \mathbf{QAS} = \mathbf{BS} \tag{2}$$

where $\mathbf{B}$ is an orthogonal matrix $(E(\mathbf{zz}^T) = \mathbf{B}E(\mathbf{SS}^T)\mathbf{B}^T = \mathbf{BB}^T = \mathbf{I})$. According to Eq. (2), the reconstructed signal can be obtained by

$$\hat{\mathbf{S}} = \mathbf{B}^T\mathbf{z} \tag{3}$$

To calculate $\mathbf{B}$, each column vector $\mathbf{b}_i$ is initialized and then updated so that ith independent component may have great non-Gaussianity (Lee et al., 2004b). Two common measures of non-Gaussianity are kurtosis and negentropy. However, the kurtosis is sensitive to outliers and hence the negentropy becomes the widely used measure of non-Gaussianity. The negentropy J of random variable y is defined as

$$J(y) = H(y_{Gaussian}) - H(y) \tag{4}$$

where $H(y) = -\int f(y)\log f(y)\,dy$ and $f(y)$ is the density of y. The $y_{Gaussian}$ is a Gaussian random variable with the same variance of y. According to Eq. (4), if $J(y)=0$, then y follows the same distribution of $y_{Gaussian}$. Thus, negentropy is non-negative and measures the departure of y from Gaussianity (Lee et al., 2004a). From Eq. (4), it is known that an estimate of probability density function is required before estimating negentropy. Thus, Hyvärinen (1999) suggested approximating negentropy by using a fixed-point algorithm for ICA (FastICA), calculated over the whitened signal $\mathbf{z}$ (i.e. Eq. (2)). In general, the FastICA calculates matrix $\mathbf{B}$ according to the following procedures.

1. Randomly choose an initial weight vector $\mathbf{b}_i$ with unit norm.
2. Let $\mathbf{b}_i \leftarrow E\{\mathbf{x}g(\mathbf{b}_i^T\mathbf{x})\} - E\{g'(\mathbf{b}_i^T\mathbf{x})\}\mathbf{b}_i$, where g is the first derivative and g' is the second derivative of G in which $G(u) = 1/a_1 \log\cosh(a_1 u)$, and a_1 is a constant and $1 \le a_1 \le 2$.
3. Normalize $\mathbf{b}_i \leftarrow \mathbf{b}_i/\|\mathbf{b}_i\|$.
4. If $\mathbf{b}_i$ has not converged, go back to Step 2.

Note that the convergence means that the dot-product of old and new values of $\mathbf{b}_i$ is equal to 1. After constructing the matrix $\mathbf{B}$, the signal can be reconstructed as $\hat{\mathbf{S}} = \mathbf{B}^T\mathbf{z} = \mathbf{B}^T\mathbf{Q}\mathbf{X} = \mathbf{W}\mathbf{X}$. The related Matlab software of FastICA toolbox can be downloaded from http://www.cis.hut.fi/projects/ica/fastica/.

ICA considers higher order statistics and tries to let components be independent. Thus, the ICA components can reveal more useful information than PCA components (Lee et al., 2004a). In the next section, the ICA-based process monitoring method will be reviewed.

2.2. ICA process monitoring

Lee et al. (2004a) proposed three measures for ICA process monitoring: I^2, I_e^2 and squared prediction error (SPE). To divide $\mathbf{W}$ into two parts: the dominant part ($\mathbf{W}_d$) and the excluded part ($\mathbf{W}_e$), the I^2 at observation k is defined as

$$I^2(k) = \hat{\mathbf{s}}_d(k)^T\hat{\mathbf{s}}_d(k) \tag{5}$$

where $\hat{\mathbf{s}}_d(k) = \mathbf{W}_d\mathbf{x}(k)$. Thus, I^2 is usually used to monitor the systematic part of process variation.

The second statistic, squared prediction error (SPE), is used to monitor the non-systematic part of common cause of variation, and it is defined as

$$SPE(k) = \mathbf{e}(k)^T\mathbf{e}(k) = (\mathbf{x}(k) - \hat{\mathbf{x}}(k))^T(\mathbf{x}(k) - \hat{\mathbf{x}}(k)) \tag{6}$$

where $\mathbf{e}(k)$ is the residual at observation k and the predictor $\hat{\mathbf{x}}(k) = \mathbf{A}\hat{\mathbf{s}}(k) = \mathbf{A}\mathbf{W}\mathbf{x}(k)$.

Another statistic, I_e^2 represents an incorrectly selected number of dominant ICA components, and it is defined as

$$I_e^2(k) = \hat{\mathbf{s}}_e(k)^T\hat{\mathbf{s}}_e(k) \tag{7}$$

where $\hat{\mathbf{s}}_e(k) = \mathbf{W}_e\mathbf{x}(k)$.

For process monitoring, the kernel density estimation (KDE) is applied to determine the control limits for I^2, SPE and I_e^2, respectively. A univariate kernel estimator with kernel K is defined by

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left\{\frac{x - x_i}{h}\right\} \tag{8}$$

where x is the considered data point, x_i is the observation, h is the smoothing parameter, n is the number of observations and K is the kernel function. There are several kernel functions adopted in the literature, among which the Gaussian kernel is the most popular one (Chen, Kruger, & Leung, 2004; Chen, Wynne, Goulding, & Sandoz, 2000; Silverman, 1986).

The aforementioned ICA process monitoring method does not take the dynamic characteristic into account. Hence, Lee et al. (2004b) suggested augmenting the observed data matrix by adding time-lagged observations and then performing FastICA algorithm. This procedure was named "dynamic ICA (DICA)". However, DICA still has some shortcomings. First, DICA is sensitive to outliers. Second, the process monitoring statistic in DICA is of an elliptical type. Therefore, an outlier rejection procedure will first be proposed. Further, a rectangular type measure is recommended to be the monitoring statistic.

3. Proposed process monitoring scheme

This section will present a novel process monitoring scheme for ICA. In the proposed approach, a measurement, namely adjusted outlyingness (AO), which is proposed by Brys, Hubert, and Rousseeuw (2005), is utilized for rejecting outliers and on-line process monitoring. The definition of AO is presented in Appendix A. Figs. 2(a) and (b) graphically illustrate the results of I^2 and AO for measuring ICA components, in which the used data are the same as in Fig. 1(c). Obviously, AO can produce a rectangular type measure, but I^2 generates an elliptical type measure.

Fig. 3 presents the framework of the proposed monitoring scheme. In outline, the proposed method includes three primary extensions. First, the original data matrix is augmented with time-lagged variables in order to take the process of autocorrelation into consideration. Second, an outlier rejection procedure is developed so as to better describe the greater part of the training data. Note that the proposed outlier rejection procedure is different from the work of Brys et al. (2005). In this study, the ICA is first performed to reduce the number of variable dimensions and then reject the outliers according to the extracted ICA components. The main advantages of this outlier rejection procedure include simplicity and shorter computation time. Third, AO is used as the monitoring statistic, and KDE is then performed to determine the control limit. The procedure of this proposed monitoring method includes off-line training and on-line process monitoring. The objective of the off-line training procedure is used to build a reference model. After that, the built model is executed on-line in order to monitor the process. The steps of off-line training are detailed as follows.

3.1. Off-line training

Step 1: Obtain a training dataset $\mathbf{X} \in \mathbf{R}^{n\times d}$ with n observations and d variables.

Step 2: Determine the time *lag* l and augment each observation vector with previous observations as in the following form:

$$X(l) = \begin{array}{c} \begin{array}{cccc} lag0 & lag1 & \cdots & lagl \end{array} \\ \left[\begin{array}{c|c|c|c} \mathbf{x}_t^T & \mathbf{x}_{t-1}^T & \cdots & \mathbf{x}_{t-l}^T \\ \mathbf{x}_{t+1}^T & \mathbf{x}_t^T & \cdots & \mathbf{x}_{t-l+1}^T \\ \vdots & \vdots & \vdots & \\ \mathbf{x}_{t+n-1}^T & \mathbf{x}_{t+n}^T & \cdots & \mathbf{x}_{t+n-1-l}^T \end{array}\right] \end{array} \tag{9}$$

where $\mathbf{x}_t^T$ denotes the d-dimensional observation vector at time t and T is the transpose operator. Ku, Storer, and Georgakis (1995) presented an iterative procedure to determine l in order to capture process dynamics. Besides, Chiang, Russell, and Braatz (2001) utilized Akaike's Information Criterion (AIC) and the subspace identification method for selecting l. Further, Lee et al. (2004b) concluded according to their experience that a value of l=1 or 2 is usually appropriate for conducting dynamic process monitoring. As mentioned above, there is no standard criterion for determining l. Thus, this study adopts l=2 to implement Step 2 according to Lees' suggestion.

Step 3: Normalize the augmented data matrix, and then perform FastICA algorithm. Thus, a demixing matrix $\mathbf{W}$ can be obtained. By selecting a few rows of $\mathbf{W}$ in which the first ith rows of $\mathbf{W}$ have the largest sum of squares (Lee et al., 2004a). The dominant part of $\mathbf{W}$ is denoted as $\mathbf{W}_d$. Hence, the m extracted independent components (ICs) can be obtained by $\hat{\mathbf{S}} = \mathbf{X}(l)\mathbf{W}_d^T$.

Step 4: Screen out outliers by AO. The rejection rule for AO is given as follows (Brys, 2005):

$$AO > Q_3 + 1.5e^{4MC}IQR \tag{10}$$

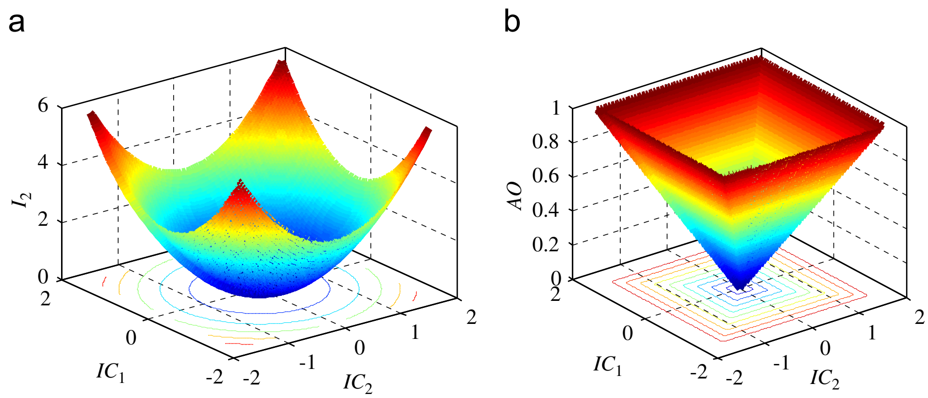

Fig. 2. I^2 and AO measures for ICA signal.

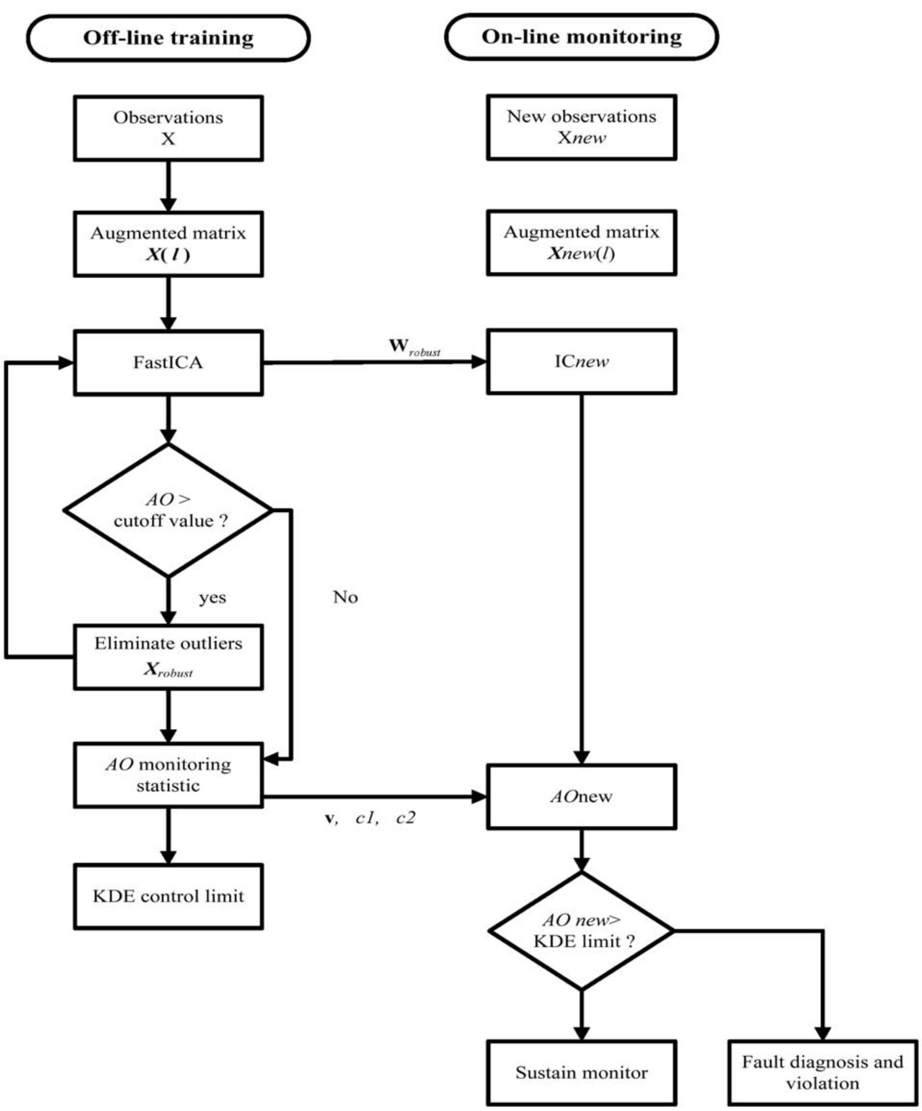

Fig. 3. The flowchart of dynamic ICA process monitoring approach.

where AO can be calculated by using Eq. (14) in Appendix A, in which the $\hat{S}$ substitutes for $\mathbf{x}'$. The MC can be obtained from Eq. (16) and $IQR = Q_3\text{-}Q_1$ is the interquartile range between Q_3 (i.e. the third quartile of projected data points $\hat{S}v^T$) and Q_1 (i.e the first quartile of projected data points $\hat{S}v^T$). Note that the filtering procedure is conducted once, in order to avoid destroying the non-Gaussianity that ICA depends on. After eliminating outliers, a robust data matrix, $\mathbf{X}_{robust}$ can be obtained. Re-run the FastICA algorithm to $\mathbf{X}_{robust}$, the robust ICs ($\hat{\mathbf{S}}_{robust}$) and dominant part of $\mathbf{W}$ ($\mathbf{W}_{robust}$) can been obtained.

Step 5: Apply $\hat{\mathbf{S}}_{robust}$ to substitute $\mathbf{x}'$ in Eq. (14) and the value of AO can be calculated. Each projection direction $\mathbf{v}$ is obtained as the normal on the hyperplane through m randomly selected data points. Further, the bound of $[c_1, c_2]$ can be calculated from Eq. (15).

Step 6: Perform the KDE method in order to determine the 99% control limit for AO measurement.

The steps of on-line process monitoring are detailed as follows.

3.2. On-line process monitoring

Step 1: Obtain a new data matrix, $\mathbf{X}_{new}$.

Step 2: Generate the augmented data matrix with *lag l* and then apply the same normalization to the augmented data matrix, denoted as $\mathbf{X}_{new}(l)$.

Step 3: Calculate ICs by $\hat{\mathbf{S}}_{new} = \mathbf{X}_{new}(l)\mathbf{W}_d^T$.

Step 4: Calculate AO measure for $\hat{\mathbf{S}}_{new}$, which is given as

$$AO_{new} = \max_{v \in H} \times \frac{[\hat{\mathbf{S}}_{new}\mathbf{v}^T\text{-}med(\hat{\mathbf{S}}_{robust}\mathbf{v}^T)]}{(c_2(\mathbf{v})\text{-}med(\hat{\mathbf{S}}_{robust}\mathbf{v}^T))I[\hat{\mathbf{S}}_{new}\mathbf{v}^T > med(\hat{\mathbf{S}}_{robust}\mathbf{v}^T)] + (med(\hat{\mathbf{S}}_{robust}\mathbf{v}^T)\text{-}c_1(\mathbf{v}))I[\hat{\mathbf{S}}_{new}\mathbf{v}^T < med(\hat{\mathbf{S}}_{robust}\mathbf{v}^T)]} \tag{11}$$

where *med* denotes *medcouple* and is defined in Eq (16). From Steps 4 and 5 of the off-line training procedure, the $\hat{\mathbf{S}}_{robust}$, $\mathbf{v}$ and $[c_1, c_2]$ can be obtained.

Step 5: Determine whether AO_{new} exceeds the control limit generated in the off-line training procedure. If an out-of-limit alarm is generated, some rectification should be enacted.

The proposed monitoring scheme takes account of the process dynamic, the contaminated dataset, and utilizes a rectangular measure for ICA. In the next section, the efficiency of this proposed method will be demonstrated through the implementation of three examples.

4. Implementation

This section first verifies the efficiency of the proposed method via a simulation example. Second, a Tennessee Eastman (TE) process is used to demonstrate the superiority of the proposed monitoring approach by comparison to several traditional methods. Finally, a real test case of a thermal power plant in Taiwan is implemented.

4.1. A simulation example

The applied simulation work is similar to Lee et al. (2004a, 2004b) and Ku (1995). Consider a dynamic process with five monitored variables as follows:

$$\mathbf{z}(k) = \begin{bmatrix} 0.118 & -0.191 & 0.287 \\ 0.847 & 0.264 & 0.943 \\ -0.333 & 0.514 & -0.217 \end{bmatrix} \mathbf{z}(k\text{-}1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \\ -2 & 1 \end{bmatrix} \mathbf{u}(k\text{-}1)$$

$$\mathbf{y}(k) = \mathbf{z}(k) + \mathbf{v}(k) \tag{12}$$

where $\mathbf{y}$ is the output with three variables (y_1, y_2, y_3). $\mathbf{v}$ is the normal distributed random vector with zero mean and variance of 0.1. $\mathbf{u}$ is the input with

$$\mathbf{u}(k) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix} \mathbf{u}(k\text{-}1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} \mathbf{w}(k\text{-}1) \tag{13}$$

$\mathbf{w}$ is a uniformly distributed random vector over interval $(-2,2)$. The input $\mathbf{u}$ and output $\mathbf{y}$, total five variables $(y_1, y_2, y_3, u_1, u_2)$ which are used for process monitoring.

A total of 1,000 observations are sampled for each simulation. The first 500 observations are used as a training dataset and the remainders are used for on-line process monitoring. A step change of w_1 by 3 is introduced at observation 500. This means that the first 500 training observations are not contaminated by outliers. In several simulation runs, the training dataset is contaminated by adding a contamination fraction (ε%) into the training dataset. In other words, there are $500 \times \varepsilon$% outliers existing in the training dataset.

For comparison purposes, several methods are also implemented, as shown below.

Scheme 1: The traditional ICA method without outlier rejection procedure and I^2 monitoring statistic is investigated by running the dataset.

Scheme 2: The DICA method without outlier rejection procedure and I^2 monitoring statistic is applied, in which two time-lagged variables are added in Eq. (9).

Scheme 3: The DICA method with Stahel–Donoho (SD) (Brys et al., 2005) outlier rejection procedure and I^2 monitoring statistic is applied, in which two time-lagged variables are added in the augmented data matrix. In short, the details of the SD method can be referred to in Brys et al. (2005), Stahel (1981), and Donoho (1982).

Scheme 4: The DICA method with AO outlier rejection procedure and I^2 monitoring statistic is applied. Also, two time-lagged variables are added in the augmented data matrix.

Scheme 5: The proposed monitoring method (i.e. DICA method with AO outlier rejection procedure and AO monitoring statistic) is investigated by using the dataset. Two time-lagged variables are also added to the augmented data matrix.

For Scheme 1, the normalization is performed to the original data matrix, whereas the normalization procedure is conducted to the augmented data matrix for Schemes 2–5. In order to make a fair comparison, the 99% control limits for all schemes are determined by the KDE method. Table 1 summarizes the process monitoring results of the above five schemes in terms of detection rate (%). Additionally, the number of outliers that were omitted from the training data for Schemes 3–5 is also listed in Table 1.

Table 1 indicates that all methods detect disturbance well when the training dataset is not contaminated. Also, the DICA methods (Schemes 2–5) perform better than the traditional ICA method (Scheme 1). Comparing the results of Schemes 1 and 2 (without outlier rejection rule) to those of Schemes 3–5 (with outlier rejection rule) shows that when the contamination is small (say $\varepsilon \le 2$%), Schemes 1 and 2 still possess satisfactory detection rates. However, when the training dataset is highly contaminated, the outlier rejection procedure can enhance the detection rate. To compare the SD and AO based rejection rule, even though Scheme 3 (SD with I^2 monitoring statistic) is comparable to Scheme 4 (AO with I^2 monitoring statistic) in terms of detection rates, the SD rejection rule discovers less outliers than AO. Further, comparing the detection rates between Schemes 4 and 5 and Table 1 presents that the AO-based monitoring statistic possesses performance superior to that of the I^2-based monitoring statistic. Thus, a rectangular measure is more suitable than the elliptical measure for describing ICA components.

4.2. Tennessee Eastman process

Several previous studies such as Ge and Song (2007), Lee et al. (2007, 2004b), and Ku et al. (1995) have used the Tennessee Eastman (TE) process as the experimental example for monitoring multivariate processes. The TE process is schematically presented in Fig. 4, which is taken from Downs and Vogel (1993). Five major units are contained in the process: an exothermic two-phase reactor, a condenser, a recycle compressor, a flash separator and a reboiler stripper. The TE process produces two products (G and H) and one by-product F from four reactants (A, C, D and E), according to the following reactions:

$A+C+D \rightarrow G$	Product 1
$A+C+E \rightarrow H$	Product 2
$A+E \rightarrow F$	By-product
$3D \rightarrow 2F$	By-product

Readers may refer to the book of Chiang et al. (2001) for a detailed description of this process. In this current study, the same simulation data generated by Chiang et al. (2001) are applied, which can be downloaded from http://brahms.scs.uiuc.edu.

Table 2 presents the 33 monitored process variables, namely 22 process measurements and 11 manipulated variables. The 21 process faults are presented in Table 3. The Fault 0 was generated with no faults, and a total of 500 observations are used as the training dataset. The testing dataset contains 960 observations, and all 21 faults are introduced at observation 160. For implementing the dynamic methods (i.e., DPCA(T^2), DICA(I^2), DICA(AO)), two time-lagged variables for each monitored variable are added. Similar to the work by Lee et al. (2004b), nine components are selected for implementing static methods (i.e., PCA(T^2), ICA(I^2), ICA(AO)) and 22 components are extracted for implementing dynamic methods. Note that the original data matrix is normalized before implementing static methods and the augmented data matrix is normalized before conducting dynamic methods. Furthermore, there are total eight outliers (observation 195, 207,224, 304, 433, 435, 446 and 488) that were omitted from the training data before conducting each monitoring scheme. For fair comparison, the 99% control limit for each process monitoring scheme is determined by KDE from Eq. (8).

The detection rates for all 21 faults under several monitoring methods are computed and tabulated in Table 4. Not all methods detected Faults 3, 9 and 15 well, since these faults are quite small and had almost no effect on the overall process (Lee et al., 2004b). On the contrary, all methods produced high detection rates for Faults 1, 2, 6, 8, 12, 13, 14 and 18.

Table 1
Comparison results in terms of detection rates (%).

Contamination on fraction (ε)	Number of outliers added	Without outlier rejection		With outlier rejection				
				SD-based rejection rule		*AO*-based rejection rule		
		Scheme 1	Scheme 2	Scheme 3	Number of omitted outliers	Scheme 4	Scheme 5	Number of omitted outliers
0%	0	96	99	99	0	99	99	0
1%	5	94	99	99	1	99	99	2
2%	10	71	98	99	5	99	99	6
3%	15	46	74	81	5	83	99	11
4%	20	37	56	69	6	71	97	12
5%	25	30	47	61	6	61	88	15

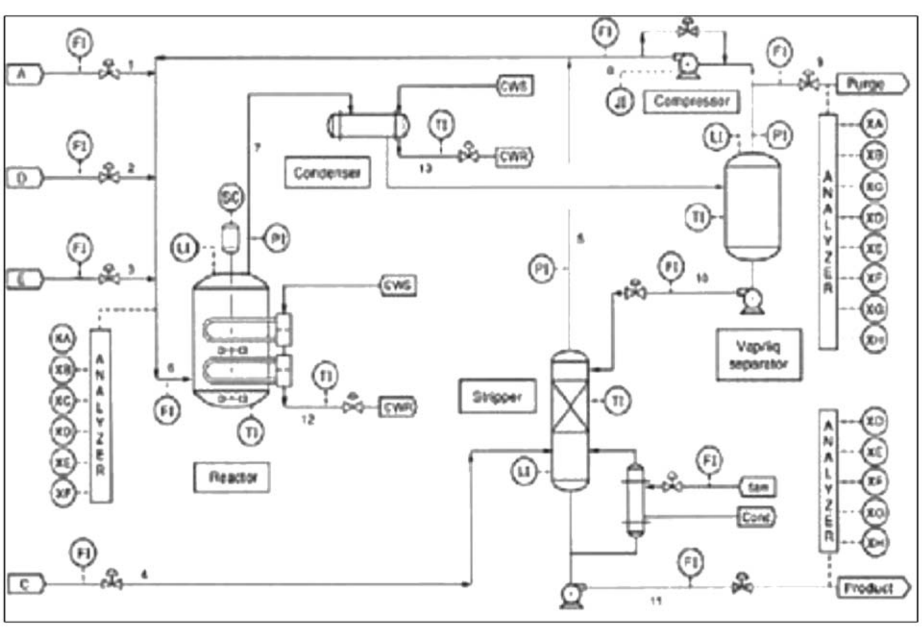

Fig. 4. Layout of TE process (Downs and Vogel, 1993).

Table 2
Monitored process variables.

No.	Process measurements	No.	Manipulated variables
1	A feed (stream 1)	23	D feed flow (stream 2)
2	D feed (stream 2)	24	E feed flow (stream 3)
3	E feed (stream 3)	25	A feed flow (stream 1)
4	A and C feed (stream 4)	26	Total feed flow valve (stream 4)
5	Recycle flow (stream 8)	27	Compressor recycle valve
6	Reactor feed rate (stream 6)	28	Purge valve (stream 9)
7	Reactor pressure	29	Separator pot liquid flow (stream 10)
8	Reactor level	30	Stripper liquid product flow (stream 11)
9	Reactor temperature	31	Stripper steam valve
10	Purge rate (stream 9)	32	Reactor cooling water valve
11	Product sep temp	33	Condenser cooling water flow
12	Product sep level		
13	Prod sep pressure		
14	Prod sep 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 temp		
22	Separator cooling water outlet temp		

Table 3
Process faults.

Fault no.	State	Disturbance
0	No fault	No
1	A/C feed ratio, B composition constant (stream 4)	Step
2	B composition, A/C 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 composition (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	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	Valve position constant (stream 4)	Constant position

Table 4
Detection rates (%) of TE process.

Faults	Static methods (nine components)			Dynamic methods (22 components)		
	PCA(T^2)	ICA(I^2)	ICA(AO)	DPCA(T^2)	DICA(I^2)	DICA(AO)
1	99	100	100	99	100	100
2	98	98	98	98	99	99
3	2	1	2	2	2	2
4	20	61	84	26	97	100
5	33	100	100	36	100	100
6	99	100	100	100	100	100
7	61	99	100	100	100	100
8	97	97	97	98	98	98
9	1	1	1	1	1	1
10	53	78	82	55	82	90
11	40	52	70	48	54	83
12	98	99	100	99	100	100
13	94	94	95	94	95	96
14	87	100	100	100	100	100
15	1	2	2	1	2	2
16	43	71	78	49	82	91
17	80	89	94	82	90	96
18	89	90	90	90	90	90
19	3	69	80	3	81	95
20	49	87	91	53	88	92
21	38	45	62	42	46	62

From Table 4 it can be seen that the ICA-based monitoring methods (ICA(I^2), ICA(AO), DICA(I^2) and DICA(AO)) outperformed PCA-based monitoring methods (PCA(T^2), DPCA(T^2)) for most fault modes. This indicates that ICA can detect non-Gaussian multivariate processes more efficiently than PCA. Fig. 5 compares the monitoring results for Fault 5 by using PCA(T^2) and ICA(I^2). Fault 5 is the step where there is a change in the condenser cooling water inlet temperature. The increased temperature will also cause a rise in the flow rate of the outlet stream from the condenser to the separator. As shown in Fig. 5, PCA(T^2) can initially detect this approximately at observation 160. However, it cannot detect the fault mode after observation 350.

Overall, the dynamic methods possess better detection rates than static methods, since the dynamic methods take process autocorrelation into account. Furthermore, the detection rates by using the rectangular monitoring statistic (AO) for ICA can outperform the elliptical based monitoring statistic (I^2). For example, DICA(AO) produces better performance than DICA(I^2) for Faults 10, 11, 16, 17 and 19. Fig. 6 illustrates the monitoring results for Fault 10 by using DICA(I^2) and DICA(AO). From Fig. 6, both methods generate no false alarms, but DICA(I^2) produces more points that fall within the control limit after observation 160 as compared to DICA(AO). In summary, the above results indicate that DICA(AO) can more efficiently monitor the process than all the other methods. Therefore it can be concluded that the proposed dynamic ICA approach can provide operators more correct information for judging the process status.

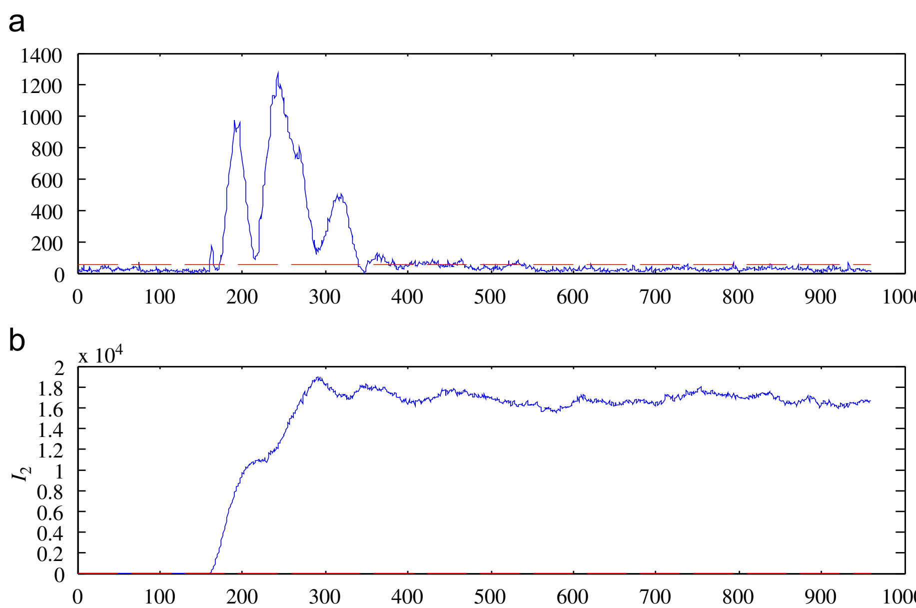

Fig. 5. Monitoring results of Fault 5: (a) PCA(T^2) and (b) ICA(I^2).

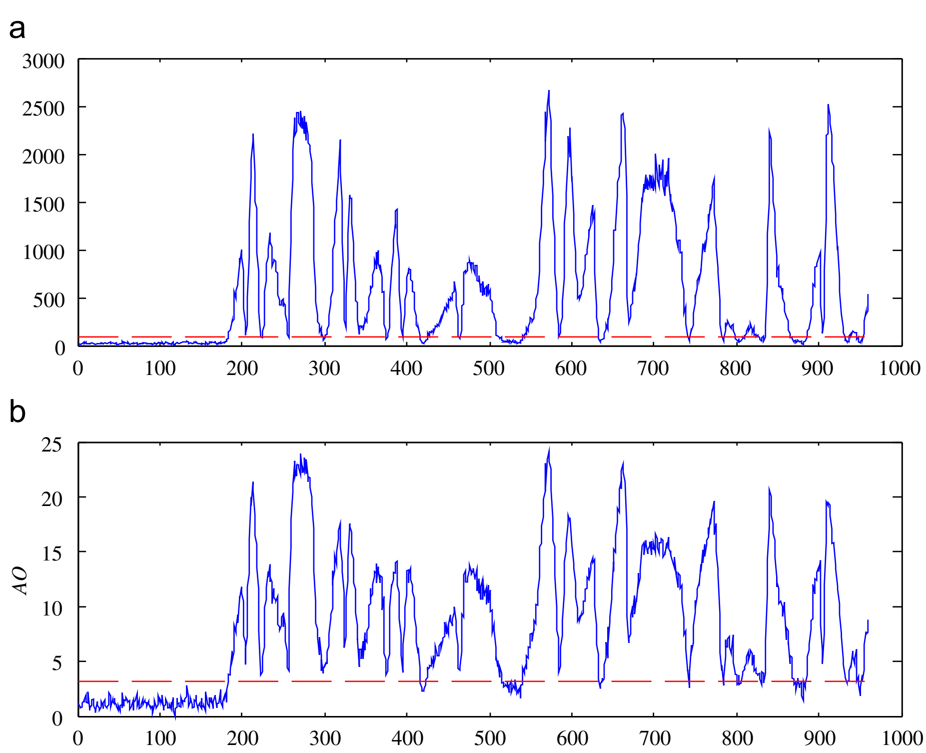

Fig. 6. Monitoring results of Fault 10: (a) DICA(T^2) and (b) DICA(AO).

4.3. *A thermal power plant case*

4.3.1. *Case description*

The power company studied herein possesses eight thermal power plants. More than 70% of the power is generated from these thermal power plants (Chien, Chen, Lo, & Lin, 2007). Thus, immediate fault detection for its equipment is an important issue. Fig. 7 shows the thermal power plant layout. Generally, the equipment in a thermal power plant consists of four major parts: the steam generator, the steam turbine generator, the electrical driven generator, and the monitoring alarm system.

1. The steam generator: The steam-generating boiler aims to produce high pressure steam required for the steam turbine that drives the electrical generator. The generator includes a boiler, water feeding system, fuel system, SCR, air heater, EP, FGD, etc.
2. The steam turbine generator: The steam turbine generator is used to transform the thermal energy into mechanical energy. The generator includes the turbine and the condensed system. It is the major piece of equipment at a thermal power plant.
3. The electrical driven generator: The electrical driven generator transforms the mechanical energy into electrical energy. The

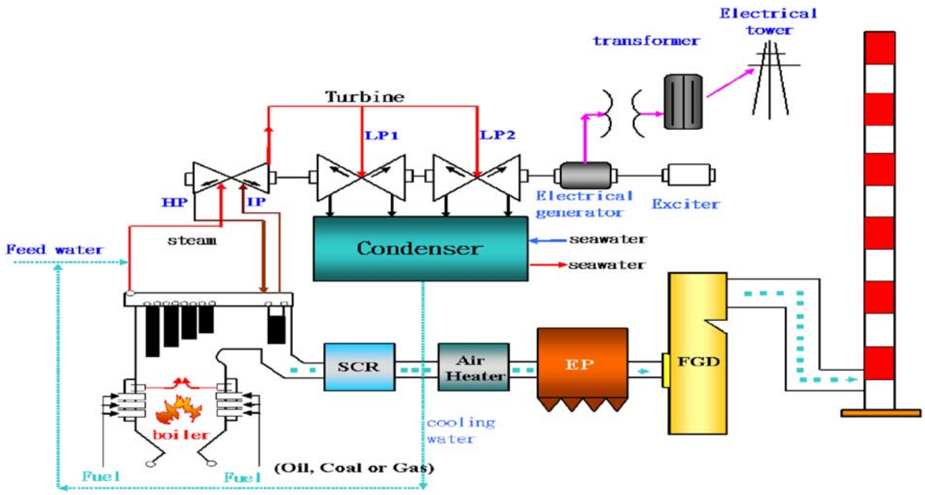

Fig. 7. Layout of thermal power plant (http://www.taipower.com.tw/).

Table 5
Abnormal parameters vs. failure modes.

Monitored parameters	Parameter variation	Failure modes
Pressure and temperature of primary steam and reheated steam	Abnormal increase	Failure in inlet steam of turbine
Pressure and temperature of primary steam and reheated steam	Abnormal decrease	Erosion of turbine blade
Vibration	Abnormal increase	Failure in bearing of turbine
Rotation speed	Over speed	Failure in blade of turbine

generator includes the electrical generator, exciter, and transformer, etc.

4. The monitoring and alarm system: This system is used to monitor the above generators, and to sound alarms if any abnormal event occurs.

Among these four systems, the steam turbine generator is the main equipment module in the thermal power plant. The key monitoring parameters of steam turbine generation include temperature and pressure of the primary steam, temperature and pressure of the reheated steam, vibration of the steam turbine generator, and rotation speed of the turbine blade. Table 5 summarizes the causes of abnormal changes in monitoring parameters that may lead to failures in steam turbine generation.

Table 6
Monitored variables.

Variable no.	Variable code	Variable name
X1	M4471	Vibration
X2	M4499	Velocity
X3	P4111	Stress
X4	P4113	Stress
X5	P4115	Stress
X6	P4120	Stress
X7	P4129	Stress
X8	P4132	Stress
X9	P4144	Stress
X10	P4145	Stress
X11	P4151	Stress
X12	T4107	Temperature
X13	T4108	Temperature
X14	T4109	Temperature
X15	T4111	Temperature
X16	T4113	Temperature
X17	T4114	Temperature
X18	T4115	Temperature
X19	T4122	Temperature
X20	T4129	Temperature
X21	T4132	Temperature
X22	T4144	Temperature
X23	T4151	Temperature
X24	T4470	Temperature
X25	T4484	Temperature
X26	T4485	Temperature
X27	T4486	Temperature
X28	T4487	Temperature
X29	T4488	Temperature

4.3.2. Implementation

The case comes from a thermal power plant which owns four 500-MW oil/gas fired units. A total of 10,960 observations were collected by the real-time monitoring system, with 29 variables monitored, which are listed in Table 6. The normality test (by using the Shapiro-Wilk statistic) for each variable is tabulated in Table 7. It indicates that all 29 variables depart from the assumption of a normal distribution. X6, X12, X18 and X24 are randomly selected to plot the autocorrelation function as shown in Fig. 8. Clearly, observations at one time are not independent over time due to the high autocorrelation in the process.

The electric power loads in the 10,960 observations are exhibited in Fig. 9. From this figure, obviously, two faults can be located. The first type of fault can be found at observations 7,098–7,675 and 10,702–10,935, in which the negative load is generated, and this abnormal situation is named Fault 1: low load. Another type of fault appears at the surrounds of peaks in Fig. 9. The effect of a high power load may rapidly increase the pressure and temperature of equipment which may cause an increase in air and water pollutions. Thus, it is also necessary to detect any high load situations, and this type of fault is named as Fault 2: overload. Analysis based on the data of electric power loading is ineffective since the faults may have occurred before detection. Hence, the proposed process monitoring approach will be applied for detecting faults by using these 29 variables.

The first 4,000 observations were used as the dataset of off-line training. The rest of the observations are used for on-line process monitoring. Six components of PCA(T^2) and ICA(I^2) are extracted for analysis. In the proposed method (DICA(AO)), two-lagged variables for each measurement are added and 10 components are extracted. The original data matrix is normalized before implementing PCA and ICA, whereas the augmented data matrix is normalized before implementing DICA. Furthermore, there are total 174 outliers that were omitted from the training data before implementing each monitoring method. In order to make a fair comparison, the KDE 99% control limit is determined for each method.

Table 7
Normality test for variables.

Variable no.	Shapiro-Wilk statistics	p-Value
X1	0.2276	<0.01
X2	0.5302	<0.01
X3	0.2036	<0.01
X4	0.1776	<0.01
X5	0.1839	<0.01
X6	0.1999	<0.01
X7	0.1767	<0.01
X8	0.1519	<0.01
X9	0.1621	<0.01
X10	0.1718	<0.01
X11	0.2809	<0.01
X12	0.4867	<0.01
X13	0.3997	<0.01
X14	0.4310	<0.01
X15	0.4485	<0.01
X16	0.5020	<0.01
X17	0.5020	<0.01
X18	0.4922	<0.01
X19	0.3572	<0.01
X20	0.3992	<0.01
X21	0.4157	<0.01
X22	0.4640	<0.01
X23	0.3971	<0.01
X24	0.3393	<0.01
X25	0.5065	<0.01
X26	0.5038	<0.01
X27	0.5051	<0.01
X28	0.4948	<0.01
X29	0.3862	<0.01

Fig. 10 shows the monitoring results of PCA(T^2), ICA(I^2) and DICA(AO). As shown in Fig. 10, PCA(T^2), ICA(I^2) and DICA(AO) can detect Fault 1: low load well. However, Fault 2 (overload) can only be discovered by ICA(I^2) and DICA(AO). Therefore, the ICA-based monitoring methods can efficiently provide information for notifying operators to reduce the rotation speed of steam turbines in order to decrease the pressure and temperature of equipment or to perform maintenance to equipment. The detection rate for ICA(I^2) is about 97% and 99% for DICA(AO). Thus, the proposed method possesses a slight superiority.

5. Conclusion

In this study, a novel dynamic process monitoring scheme for ICA has been developed and presented. The advantage of this proposed method takes the process dynamic into consideration. Further, the proposed AO outlier rejection procedure has been shown to eliminate outliers before implementing ICA. Additionally, a rectangular type measure, AO, was used as the monitoring statistic. Through investigating a five-variable simulation example, the rejection procedure was seen to be more efficient when the training dataset was contaminated. The TE process demonstrated that the proposed monitoring method possessed superior performance for most faults in comparison to other alternatives. Finally, a real case of a thermal power plant showed that the proposed method can correctly detect fault types when compared to PCA.

The proposed method can be extended to other ICA algorithms, such as kernel ICA, multiway ICA and so forth. Determining the number of lags (l) is also an important issue worthy of further

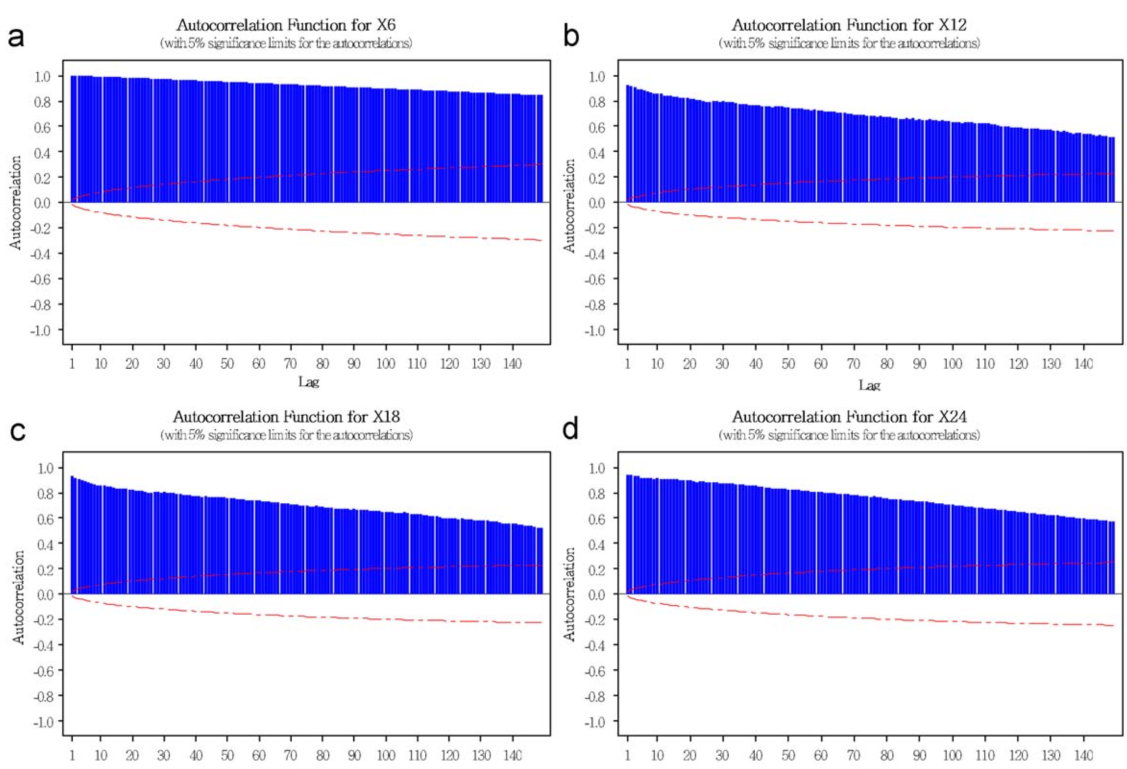

Fig. 8. Autocorrelation function plots for: (a) X6; (b) X12; (c) X18; and (d) X24.

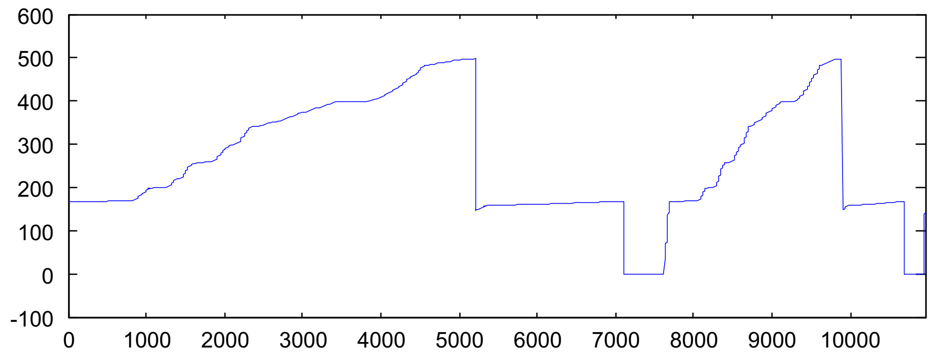

Fig. 9. The electric power loads.

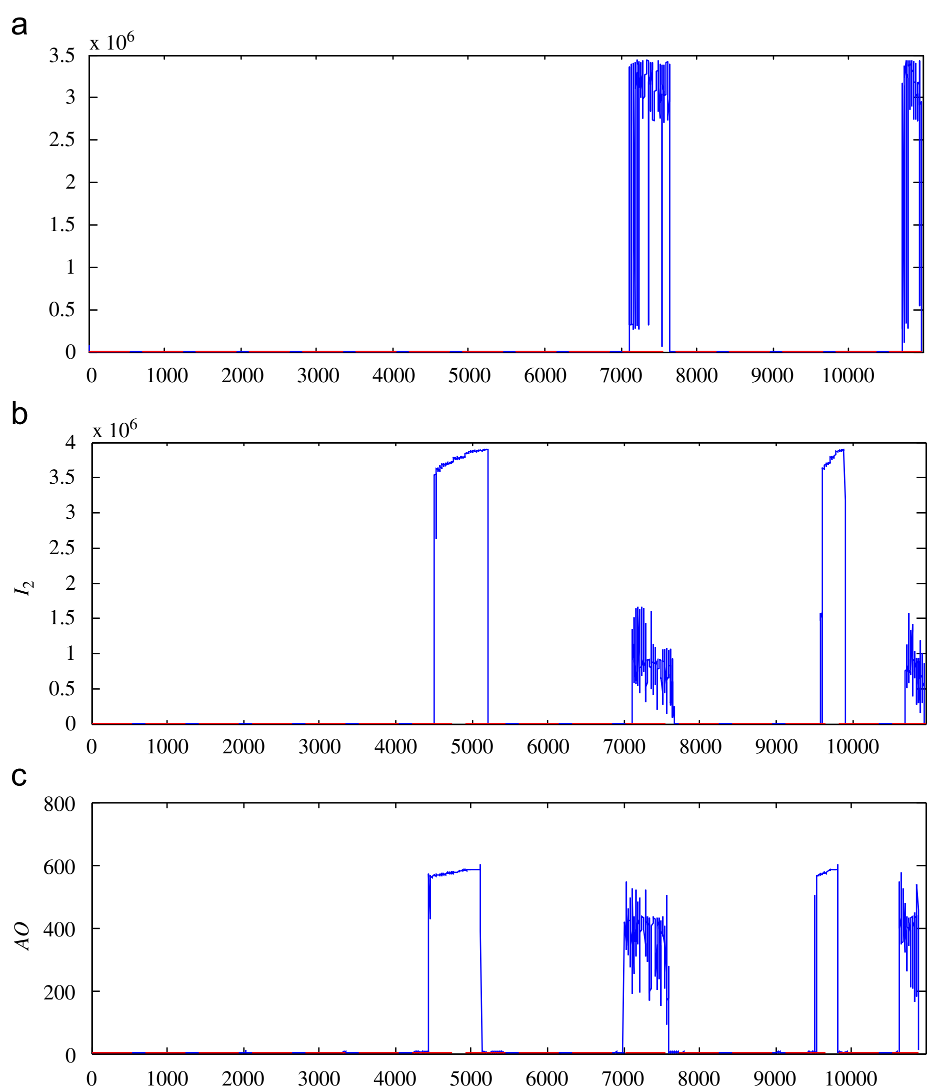

Fig. 10. The monitoring results of thermal power plant: (a) PCA(T^2); (b) ICA(I^2); and (c) DICA(AO).

investigation. Last but not least, reduction in computational time of *AO* is an additional benefit.

Acknowledgment

This work was supported in part by National Science Council of Taiwan (Grant no. NSC 98-2410-H-324-006-MY2).

Appendix A

The adjusted outlyingness (AO) is defined as

$$AO_i = \max_{v \in H} \times \frac{[\mathbf{x}_i' \mathbf{v} - med(\mathbf{x}_j' \mathbf{v})]}{(c_2(\mathbf{v}) - med(\mathbf{x}_j' \mathbf{v}))I[\mathbf{x}_i' \mathbf{v} > med(\mathbf{x}_j' \mathbf{v})] + (med(\mathbf{x}_j' \mathbf{v}) - c_1(\mathbf{v}))I[\mathbf{x}_i' \mathbf{v} < med(\mathbf{x}_j' \mathbf{v})]} \tag{14}$$

where $\mathbf{x}$ represents the observed data vector, *med* denotes median operator and $\mathbf{v}$ is the projection direction. The bound of c_1 and c_2

can be obtained by

$$
\begin{aligned}
&[c_1, c_2] = [Q_1\text{-}1.5e^{-4MC}IQR, Q_3 + 1.5e^{3MC}IQR] \quad \text{if } MC > 0 \\
&[c_1, c_2] = [Q_1\text{-}1.5e^{-3MC}IQR, Q_3 + 1.5e^{4MC}IQR] \quad \text{if } MC < 0
\end{aligned} \tag{15}
$$

where Q_1 and Q_3 are the first and third quartiles, respectively, of the projected data $\mathbf{x}_i' \mathbf{v}$, and $IQR = Q_3\text{-}Q_1$. The *MC* means *medcouple* (Brys, Hubert, and Strufy, 2004) and is a robust measure of skewness which is given as

$$
MC(g_1, \ldots, g_n) = \underset{i,j}{med} \frac{(g_j\text{-}med_k g_k)\text{-}(g_i\text{-}med_k g_k)}{g_j\text{-}g_i} \tag{16}
$$

where i and j satisfy $g_i \le med_k(g_k) \le g_j$and $g_i \ne g_j$. To implement the *AO* measure, readers may download the software of LIBRA Matlab toolbox from http://www.wis.kuleuven.ac.be/stat/robust.html, and the user guide can be obtained from Verboven and Hubert (2005).

References

Available at: 〈http://www.taipower.com.tw/〉.

Brys, G., Hubert, M., & Rousseeuw, P. J. (2005). A robustification of independent component analysis. *Journal of Chemometrics*, *19*, 364–375.

Brys, G., Hubert, M., & Strufy, A. (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics*, *13*, 996–1017.

Chen, Q., Kruger, U., & Leung, A. T.Y. (2004). Regularised kernel density estimation for clustered process data. *Control Engineering Practice*, *12*, 267–274.

Chen, Q., Wynne, R. J., Goulding, P., & Sandoz, D. (2000). The application of principal component analysis and kernel density estimation to enhance process monitoring. *Control Engineering Practice*, *8*, 531–543.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems*. London: Springer.

Chien, C. F., Chen, W. C., Lo, F. Y., & Lin, Y. C. (2007). A case study to evaluate the productivity changes of the thermal power plants of the Taiwan power company. *IEEE Transactions on Energy Conversion*, *22*, 680–688.

Donoho, D. L. (1982). *Breakdown properties of multivariate location estimators*. Ph.D. qualifying paper, Harvard University.

Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers and Chemical Engineering*, *17*(3), 245–255.

Ge, Z., & Song, Z. (2007). Process monitoring based on independent component analysis—principal component analysis (ICA-PCA) and similarity factors. *Industrial and Engineering Chemistry Research*, *46*, 2054–2063.

Ge, Z., & Song, Z. (2008). Online monitoring of nonlinear multiple mode processes based on adaptive local model approach. *Control Engineering Practice*, *16*, 1427–1437.

González, I., & Sánchez, S. (2007). Principal alarms in multivariate statistical process control using independent component analysis. *International Journal of Production Research*, *46*(22), 6345–6366.

Hubert, M., & Van der Veeken, S. (2008). Outlier detection for skewed data. *Journal of Chemometrics*, *22*, 235–246.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent analysis. *IEEE Transactions on Neural Networks*, *10*, 626–634.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., & Ohno, H. (2003). Monitoring independent components for fault detection. *A.I.Ch.E. Journal*, *49*(4), 969–976.

Ku, W., Storer, R. H., & Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *30*, 179–196.

Lee, J. M., Qin, S. J., & Lee, I. B. (2006). Fault detection and diagnosis based on modified independent component analysis. *A.I.Ch.E. Journal*, *52*(10), 3501–3514.

Lee, J. M., Qin, S. J., & Lee, I. B. (2007). Fault detection of non-linear process using kernel independent component analysis. *The Canadian Journal of Chemical Engineering*, *85*, 526–536.

Lee, J. M., Yoo, C. K., & Lee, I. B. (2004a). Statistical process monitoring with independent component analysis. *Journal of Process Control*, *14*, 467–485.

Lee, J. M., Yoo, C. K., & Lee, I. B. (2004b). Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science*, *59*, 2995–3006.

Lu, C. J., Wu, C. M., Keng, C. J., & Chiu, C. C. (2008). Integrated application of SPC/EPC/ICA and neural networks. *International Journal of Production Research*, *46*(4), 873–893.

Martin, E. B., & Morris, A. J. (1996). Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, *6*(6), 349–358.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. UK: Chapman & Hall.

Stahel, W. A. (1981). *Robust estimation: Infinitisimal optimality and covariance matrix estimators*. Ph.D. thesis, ETH Zürich.

Verboven, S., & Hubert, M. (2005). LIBRA: A MATLAB library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, *75*, 127–136.

Xia, C., & Howell, J. (2004). Isolating multiple sources of plant-wide oscillations via independent component analysis. *Control Engineering Practice*, *13*, 1027–1035.

Yoo, C. K., Lee, J. M., Vanrolleghem, P. A., & Lee, I. B. (2004). On-line monitoring of batch processes using multiway independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, *71*(2), 151–163.

Zhu, K., Hong, G. S., Wong, Y. S., & Wang, W. (2008). Cutting force denoising in micro-milling tool condition monitoring. *International Journal of Production Research*, *46*(16), 4391–4408.