Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Minimum mean squared prediction error criterion based improved independent component analysis method for process monitoring

Lianfang Cai, Xuemin Tian, Yuping Cao

College of Information and Control Engineering, China University of Petroleum, Qingdao, 266580, China (E-mail: cailianfang@163.com, tianxm@upc.edu.cn, cao_yp@163.com).

Abstract: Independent component analysis (ICA) is a newly emerging feature extraction method for non-Gaussian process monitoring. However, the extracted feature by ICA may not represent the original process data well, which can result in the degraded monitoring performance. In this paper, an improved ICA method based on the minimum mean squared prediction error criterion is proposed for process monitoring. A new criterion which can make the extracted non-Gaussian feature be efficient representation for the original process data is constructed as the objective function of the improved ICA by integrating the maximum negentropy criterion of the conventional ICA with the minimum mean squared prediction error criterion. Then the gradient ascent algorithm is applied to optimize the constructed objective function for seeking the feature extraction directions. Finally, a monitoring statistic is built based on the extracted feature to detect process faults. The simulation studies on the Tennessee Eastman benchmark process demonstrate that the improve ICA is more effective than the conventional ICA for improving the monitoring performance in terms of the fault detection rate.

Keywords: Process monitoring, independent component analysis, feature extraction, non-Gaussian, minimum mean squared prediction error, fault detection rate

1. INTRODUCTION

As industrial processes become more complex, process monitoring plays a critical role in ensuring the production safety and improving the product quality. Multivariate statistical process monitoring (MSPM), which is a kind of data-driven methods, has drawn much attention from academic researchers and process engineers (Jiang and Yan, 2012; Stubbs et al., 2012; Yu, 2011; Zhang et al., 2009). Among the numerous MSPM methods, principal component analysis (PCA) is a classical approach and has a wide range of applications in process monitoring (BinShams et al., 2011; Jia et al., 2010; Shao et al., 2009; Wang and Yuan, 2007; Yu, 2012). It projects the original process variables onto a smaller set of uncorrelated principal components (PCs) which can retain most of the original variance. However, PCA can only consider up to second-order statistics and can not make use of higher-order statistical information contained in process data (Lee et al., 2006), which may lead to inadequate feature extraction and degraded monitoring performance.

Independent component analysis (ICA), originated from the blind source separation problem, is recently introduced into the process monitoring field (Ge et al., 2012; Hsu et al., 2010; Lee et al., 2011; Stefatos and Hamza, 2010; Zhang et al., 2010; Zhao et al., 2008). Different from PCA, ICA can utilize the higher-order statistical information of process data and extract non-Gaussian feature called independent components (ICs) from original process variables. Compared with PCs extracted by PCA, ICs can reveal more useful process information and realize the process monitoring more precisely (Wang et al., 2012). Many extensions of ICA have been developed to account for different process characteristics. Tian et al. (2009) proposed a multiway kernel ICA method based on feature samples for nonlinear batch process monitoring. Considering the process dynamics, Odiowei et al. (2010) proposed a state-space ICA based process monitoring method. Rashid et al. (2012) proposed a hidden Markov model based adaptive ICA method for monitoring complex chemical process with multiple operating conditions and inherent system uncertainty. Taking both the nonlinearity and multimodality of process into account, Zhang et al. (2013) proposed a modified kernel ICA based monitoring method. In these studies, ICA has been thought as a promising feature extraction and dimension reduction method. However, the maximum negentropy criterion in conventional ICA only concentrates on extracting the non-Gaussian feature from the process data whereas ignores to make the extracted feature represent the process data as well as possible, which can result in the insufficient feature extraction and thus the adverse effect on the monitoring performance.

Motivated by the above analysis, we propose an improved ICA method based on minimum mean squared prediction error criterion for process monitoring in this paper. The maximum negentropy criterion commonly used in conventional ICA is integrated with the minimum mean squared prediction error criterion, and an improved ICA objective function which can make the extracted non-Gaussian feature represent the original process data sufficiently is constructed. Subsequently, the built objective function is optimized by the gradient ascent algorithm to obtain the feature extraction directions. Lastly, a monitoring

statistic is established to conduct process monitoring based on the extracted feature. The remainder of this paper is organized as follows. In section 2, the principle of conventional ICA is described. Section 3 proposes an improved ICA method based on the minimum mean squared prediction error criterion. Section 4 formulates the process monitoring strategy using the improved ICA method. Simulation studies on the Tennessee Eastman benchmark process are investigated and discussed in section 5. Finally, conclusions are drawn in section 6.

2. THE PRINCIPLE OF CONVENTIONAL ICA

In this section, the principle of the original ICA will be simply reviewed, which will be the basis of the proposed improved ICA method in this paper.

The measured variables $\boldsymbol{x}=[x_1,x_2,\cdots,x_m]^{\mathrm{T}}\in R^{m\times 1}$ are supposed to be linear combinations of the unknown zero-mean ICs $\boldsymbol{s}=[s_1,s_2,\cdots,s_m]^{\mathrm{T}}\in R^{m\times 1}$. The relationship between them can be given by

$$\boldsymbol{x}=\boldsymbol{A}\boldsymbol{s} \tag{1}$$

where $\boldsymbol{A}\in R^{m\times m}$ is the unknown mixing matrix.

The ICA tries to estimate both $\boldsymbol{A}$ and $\boldsymbol{s}$ only from the known $\boldsymbol{x}$. Alternatively, the objective of the ICA can be defined as follows: to find a de-mixing matrix $\boldsymbol{W}\in R^{m\times m}$ which can make the elements of the estimated vector $\hat{\boldsymbol{s}}\in R^{m\times 1}$ given by

$$\hat{\boldsymbol{s}}=\left[\hat{s}_1,\hat{s}_2,\cdots,\hat{s}_m\right]^{\mathrm{T}}=\boldsymbol{W}\boldsymbol{x} \tag{2}$$

be as independent of each other as possible.

Usually, the process variables need to be whitened before applying the ICA algorithm. The whitened variables $\boldsymbol{z}\in R^{m\times 1}$ can be obtained as

$$\boldsymbol{z}=\boldsymbol{Q}\boldsymbol{x} \tag{3}$$

where $\boldsymbol{Q}=\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}^{\mathrm{T}}\in R^{m\times m}$ denotes the whitening matrix, $\boldsymbol{\Lambda}=diag(\lambda_1,\lambda_2,\cdots,\lambda_m)\in R^{m\times m}$, λ_i, $i=1,2,\cdots,m$ are the eigenvalues of $\boldsymbol{C}_x=E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\in R^{m\times m}$ and satisfy the condition $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_m$, $\boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_m]\in R^{m\times m}$ is an orthogonal matrix, $\boldsymbol{v}_i$, $i=1,2,\cdots,m$ are the eigenvectors of $\boldsymbol{C}_x$, $E[\cdot]$ denotes the expectation operator.

The whitened variables satisfy the condition

$$E[\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}]=\mathbf{I}_m \tag{4}$$

where $\mathbf{I}_m\in R^{m\times m}$ is the identity matrix.

For the mathematical convenience, all the ICs can be assumed to have the unit variance without loss of generality. Then (2) can be reformulated as

$$\hat{\boldsymbol{s}}=\boldsymbol{W}\boldsymbol{x}=\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}=\boldsymbol{U}\boldsymbol{z} \tag{5}$$

where $\boldsymbol{W}=\boldsymbol{U}\boldsymbol{Q}$, $\boldsymbol{U}=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_m]^{\mathrm{T}}\in R^{m\times m}$ is an orthogonal matrix due to the reason which is $E[\hat{\boldsymbol{s}}\hat{\boldsymbol{s}}^{\mathrm{T}}]=E[\boldsymbol{U}\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}]=\boldsymbol{U}E[\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}]\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\mathbf{I}_m$.

To calculate the orthogonal matrix $\boldsymbol{U}$, the optimization objective of the ICA can be defined as follows (Hyvarinen, 1999)

$$\begin{aligned}&\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}} J_1(\boldsymbol{u}^{\mathrm{T}})=\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E[G(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z})]-E[G(\upsilon)]\}^2\\&\text{s.t. } E[(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z})^2]=1,\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1\end{aligned} \tag{6}$$

where $\boldsymbol{u}^{\mathrm{T}}\in R^{1\times m}$ is a row vector of the orthogonal matrix $\boldsymbol{U}$, υ is a Gaussian variable with zero mean and unit variance, $G(\cdot)$ is a non-quadratic function. The specific details of the ICA can be found in (Hyvarinen, 1999).

Once the orthogonal matrix $\boldsymbol{U}$ is obtained, the ICs can be estimated by using (5).

3. IMPROVED ICA BASED ON MINIMUM MEAN SQUARED PREDICTION ERROR

Although the estimated ICs $\hat{\boldsymbol{s}}$ by the above conventional ICA can reveal important statistical information from process data, they may still run the risk of not representing the process data well. The reason is that the above ICA does not take the minimum mean squared prediction error criterion which is widely used for feature extraction and dimension reduction into account (Wang et al., 2012). As a result, the estimated ICs may not reconstruct the original process data well and some important process information may be lost. In this section, we propose an improved ICA method based on the minimum mean squared prediction error criterion to solve such the problem of the above ICA.

As the relationship of the vector $\boldsymbol{u}^{\mathrm{T}}$ and the corresponding estimated IC $\hat{s}\in R$ can be written as $\hat{s}=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}$, the mean squared prediction error (MSPE) for the vector $\boldsymbol{u}^{\mathrm{T}}$ and the extracted IC $\hat{s}$ can be defined as

$$\begin{aligned}&\mathrm{MSPE}=E[(\boldsymbol{x}-\boldsymbol{Q}^{-1}\boldsymbol{u}\hat{s})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{Q}^{-1}\boldsymbol{u}\hat{s})]\\&=E[(\boldsymbol{x}-\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})]\\&\text{s.t. } \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1\end{aligned} \tag{7}$$

From (7), it is well acknowledged that if the vector $\boldsymbol{u}^{\mathrm{T}}$ and the extracted feature $\hat{s}$ can well reconstruct the original process data, they should make the MSPE as small as possible. Based on this analysis, we take the minimum MSPE as the criterion for seeking the vector $\boldsymbol{u}^{\mathrm{T}}$ and further derive (7) as follows, after which the relationship of the minimum MSPE criterion and the vector $\boldsymbol{u}^{\mathrm{T}}$ can be found more obviously.

$$\begin{aligned}
&\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\mathrm{MSPE}=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}E[\boldsymbol{x}^{\mathrm{T}}(\mathbf{I}_m-\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}})(\mathbf{I}_m-\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{x}]\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}E[\boldsymbol{x}^{\mathrm{T}}(\mathbf{I}_m-\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}-\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\\
&\qquad+\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{x}]\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})-E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\boldsymbol{x})\\
&\qquad+E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})\}\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-E(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u})-E(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u})\\
&\qquad+E(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u})\}\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}E(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{Q}^{-1}\boldsymbol{u}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}E(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\\
&\qquad+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}^{-\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}E(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{u}\}\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{u}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{u}\\
&\qquad+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{u}\}\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{u}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{u}+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{u}\}\\
&=\min_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{u}\} \qquad (8)
\end{aligned}$$

From (8), we can find that the minimum MSPE criterion only depends on the vector $\boldsymbol{u}^{\mathrm{T}}$ since the first item $E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})$ is a constant. Then, the minimum MSPE criterion can be transformed into the following equivalent form

$$\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}} J_2(\boldsymbol{u}^{\mathrm{T}})=\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{u}\quad \text{s.t. } \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1 \qquad (9)$$

In order to make the extracted non-Gaussian feature represent the process data effectively, we can integrate (9) with the maximum negentropy criterion of (6) and construct a new objective function as follows

$$\begin{aligned}
&\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}} J(\boldsymbol{u}^{\mathrm{T}})=\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{\eta J_1(\boldsymbol{u}^{\mathrm{T}})+(1-\eta)J_2(\boldsymbol{u}^{\mathrm{T}})\}\\
&=\max_{\boldsymbol{u}^{\mathrm{T}}\in\boldsymbol{U}}\{\eta[E(G(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z}))-E(G(\upsilon))]^2+(1-\eta)\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{u}\} \qquad (10)\\
&\text{s.t. } E[(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z})^2]=1,\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}=1
\end{aligned}$$

where η is a predefined weight used to place different values on the two subobjective functions according to the practical necessity and satisfies the condition $0\le\eta\le 1$, the non-quadratic function $G(\cdot)$ can be chosen as $G(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z})=-\exp(-(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{z})^2/2)$ (Hyvarinen, 1999).

The gradient ascent algorithm can be adopted to optimize (10) as

$$\begin{aligned}
&\boldsymbol{u}(k+1)=\boldsymbol{u}(k)+\mu\frac{\partial J(\boldsymbol{u}^{\mathrm{T}})}{\partial\boldsymbol{u}}|_k\\
&=\boldsymbol{u}(k)+\mu\{2\eta[E(G(\boldsymbol{u}(k)^{\mathrm{T}}\boldsymbol{z}))-E(G(\upsilon))]E(\boldsymbol{z}g(\boldsymbol{u}(k)^{\mathrm{T}}\boldsymbol{z})) \qquad (11)\\
&\quad+2(1-\eta)\boldsymbol{\Lambda}\boldsymbol{u}\}
\end{aligned}$$

where $\boldsymbol{u}(k)$ denotes the value of $\boldsymbol{u}$ at the k th iteration, μ is the step size parameter, the function $g(\cdot)$ is the first order derivative of $G(\cdot)$.

On the basis of (11), we develop an improved ICA algorithm for calculating the orthogonal matrix $\boldsymbol{U}$ as follows

1) Let counter $i=1$.

2) Set an initial value with unit norm for $\boldsymbol{u}_i$.

3) Update $\boldsymbol{u}_i$ by using (11).

4) Do the following orthogonalization and normalization: $\boldsymbol{u}_i=\boldsymbol{u}_i-\sum_{j=1}^{i-1}(\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j)\boldsymbol{u}_j$, $\boldsymbol{u}_i=\boldsymbol{u}_i/\|\boldsymbol{u}_i\|_2$.

5) If $\boldsymbol{u}_i$ has not converged, go back to 3); else, save the vector $\boldsymbol{u}_i$ and go to the next step.

6) Let $i=i+1$. If $i\le m$, go back to 2); else, stop the procedure.

Once the orthogonal matrix $\boldsymbol{U}$ is obtained, we can also estimate the ICs $\hat{\boldsymbol{s}}$ by using (5).

4. PROCESS MONITORING BASED ON IMPROVED ICA

In order to detect process faults, the estimated ICs by the proposed ICA method should be arranged in the descending order according to their importance and the row vectors of the orthogonal matrix $\boldsymbol{U}$ should also be sorted correspondingly. The d estimated dominant ICs $\hat{\boldsymbol{s}}_d\in R^{d\times 1}$ can then be expressed as

$$\hat{\boldsymbol{s}}_d=\boldsymbol{U}_d\boldsymbol{z}=[\hat{s}_1,\hat{s}_2,\cdots,\hat{s}_d]^{\mathrm{T}} \qquad (12)$$

where $\boldsymbol{U}_d=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_d]^{\mathrm{T}}$ and $d<m$.

Based on the extracted feature, a monitoring statistic can be defined as

$$I^2=\hat{\boldsymbol{s}}_d{}^{\mathrm{T}}\hat{\boldsymbol{s}}_d=(\boldsymbol{U}_d\boldsymbol{z})^{\mathrm{T}}(\boldsymbol{U}_d\boldsymbol{z})=(\boldsymbol{U}_d\boldsymbol{Q}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{U}_d\boldsymbol{Q}\boldsymbol{x}) \qquad (13)$$

The δ confidence limit for the I^2 monitoring statistic can be determined according to (Russell et al., 2000). Split the normal process data into two parts: the training dataset with sample size N_1 and the validating dataset with sample size N_2. Based on the training data, calculate the whitening matrix $\boldsymbol{Q}$ and the matrix $\boldsymbol{U}_d$. Based on the validating data, calculate the time series of the I^2 statistic by (13). Then round $N_2(1-\delta)$ toward the nearest integer r. Adopt the r th highest value of the time series of the I^2 statistic as the δ confidence limit.

The monitoring strategy based on the improved ICA can be divided into two stages: offline modeling stage and online fault detection stage. The details can be described as follows

Stage I: offline modeling stage

1) Divide the normal process data into two parts: the training dataset with sample size N_1 and the validating dataset with sample size N_2.

2) Calculate the mean and variance of each process variable in the training dataset and scale both the training data and validating data using the obtained means and variances.

3) Based on the training data, calculate the whitening matrix $\boldsymbol{Q}$ and the matrix $\boldsymbol{U}_d$.

4) Based on the validating data, calculate the time series of the I^2 statistic by (13).

5) Determine the δ confidence limit.

Stage II: online fault detection stage

1) Scale the current process data with the means and variances of the training data.

2) Calculate the current value of the I^2 monitoring statistic by (13).

3) Determine whether the current statistic value exceeds the confidence limit and give an alarm if some fault is detected.

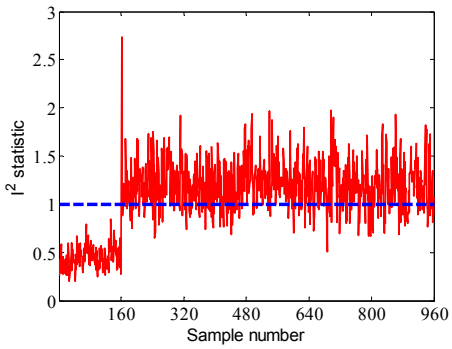
5. SIMULATION STDUIES

The Tennessee Eastman (TE) process is a well-known benchmark for testing the performance of various process monitoring methods (Ge et al., 2012; Lee et al., 2011; Mahadevan and Shah, 2009; Stefatos and Hamza, 2010; Yu, 2011; Yu, 2012; Zhang et al., 2010). It consists of five major units: a reactor, a condenser, a compressor, a separator, and a stripper. The four reactants A, C, D, E and the inert B are fed into the reactor where the products G and H are formed and a byproduct F is also produced. The process has 52 monitoring variables and allows 21 pre-programmed process faults. There are 960 observations for normal and each fault condition with a sampling interval of 3 minutes. Each fault is introduced at the 160th sample. More details about the TE process can be found in (Chiang et al., 2001).

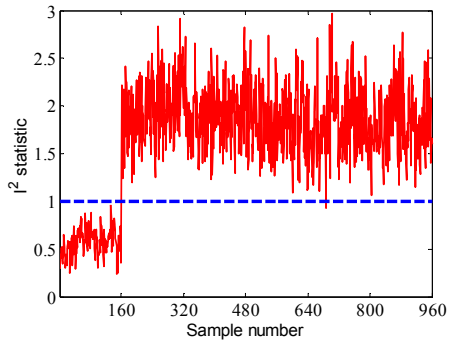
In this section, the process monitoring performance of the conventional ICA and improved ICA is investigated with the TE process. The performance is evaluated from the fault detection rate which is defined as the percentage of the alarming fault samples in all the fault samples. The normal dataset is divided into two parts: the training dataset with size 500 to build the monitoring model and the validating dataset with size 460 to calculate the 99% confidence limit. The parameter η in (10) is set as 0.5 to balance the two subobjective functions. The ICs are ordered according to the L_2 norm of the row vectors of the de-mixing matrix (Lee et al., 2004). The number d of dominant ICs for both methods is chosen as 30 so that the eigenvalues cumulative sum of the covariance matrix $\boldsymbol{C}_x$ is above 90%.

The monitoring charts of the fault 4 are firstly illustrated in Fig. 1 to show the effectiveness of the proposed method. The monitoring statistic values are plotted as the solid line and the confidence limit is plotted as the dashed line. For convenience of comparisons, in each monitoring chart, all monitoring statistic values are divided by their corresponding confidence limit and then the confidence limit is equal to one.

It can be seen from the monitoring chart of the improved ICA that, after the occurrence of the fault 4 at the 160th sample, the monitoring statistic values exceed the confidence limit obviously and almost all the fault samples are detected successfully. However, in the monitoring chart of the conventional ICA, the monitoring statistic values of many fault samples are still below the confidence limit and thus can not indicate the fault effectively and confidently. Thus, the monitoring charts comparison suggests that the improved ICA is more sensitive to the fault 4 than the conventional ICA. The monitoring charts of fault 19 are also used for methods comparison. Due to the proprietary reason, the information of this fault is unknown and only the fault data are provided. The monitoring charts on this fault are shown in Fig. 2. It is obvious that there are much more fault samples whose monitoring statistic values exceed the confidence limit in the monitoring chart of the improved ICA, leading to a much higher fault detection rate than the conventional ICA.

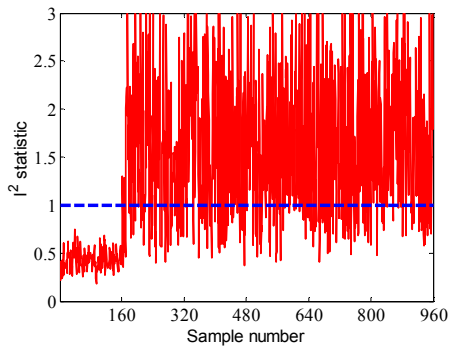


(a) The monitoring chart of the conventional ICA

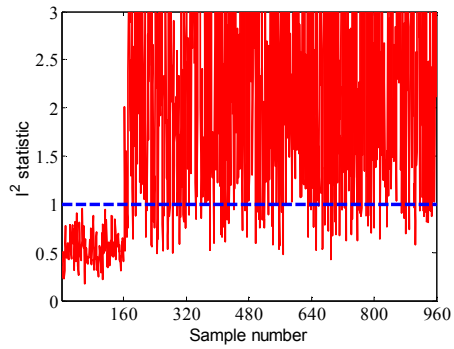


(b) The monitoring chart of the improved ICA

Fig. 1. The monitoring charts of the two methods on fault 4



(a) The monitoring chart of the conventional ICA



(b) The monitoring chart of the improved ICA

Fig. 2. The monitoring charts of the two methods on fault 19

The fault detection rate of the two methods on all the 18 fault cases of the TE process is tabulated in the Table 1. The faults 3, 9, and 15 are excluded, because there are no observable changes in the mean or the variance of those three fault datasets and these three faults have been proven to be difficult for data-driven methods (Zhang et al., 2011). Obviously, for the faults 1, 2, 5-8, 12-14, and 18, the two methods have similar fault detection rate. However, for the faults 4, 10, 11, 16, 17, and 19-21, the improved ICA achieves much higher fault detection rate than the conventional ICA. The average fault detection rate of the conventional ICA is 0.8766, whereas the average fault detection rate of the improved ICA is 0.9214, which also shows the superior fault detection ability of the improved ICA. According to the above analysis, it can be concluded that the fault detection performance of the improved ICA is much better than that of the conventional ICA.

Table 1. The fault detection rate comparison

Fault	ICA-I^2	Improved ICA-I^2
1	0.9975	0.9975
2	0.9850	0.9850
4	0.8025	0.9988
5	1.0000	1.0000
6	1.0000	1.0000
7	1.0000	1.0000
8	0.9762	0.9788
10	0.8225	0.8788
11	0.5362	0.6800
12	0.9962	0.9975
13	0.9475	0.9538
14	0.9988	0.9988
16	0.8150	0.8988
17	0.9188	0.9550
18	0.8962	0.9012
19	0.7512	0.8400
20	0.8275	0.9100
21	0.5075	0.6112

6. CONCLUSIONS

This paper proposed an improved ICA method for process monitoring. The proposed method can not only extract the important non-Gaussian information from the original process data but also make the extracted feature represent the original data efficiently and sufficiently by integrating the maximum negentropy criterion and the minimum mean squared prediction error criterion. The simulation example of the TE process demonstrates that the proposed method can effectively detect process faults and its process monitoring performance outperforms the conventional ICA.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (61273160), the Fundamental Research Funds for the Central Universities (12CX06071A), the Doctoral Fund of Shandong Province (BS2012ZZ011) and the Postgraduate Innovation Funds of China University of Petroleum (CX2013060).

REFERENCES

BinShams, M.A., Budman, H.M., and Duever, T.A. (2011). Fault detection, identification and diagnosis using CUSUM based PCA. *Chemical Engineering Science*, **66(20)**, 4488-4498.

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

Ge, Z.Q., Xie, L., Kruger, U., and Song, Z.H. (2012). Local ICA for multivariate statistical fault diagnosis in systems with unknown signal and error distribution. *AIChE*, **58(8)**, 2357-2372.

Hsu, C.C., Chen, L.S., and Liu, C.H. (2010). A process monitoring scheme based on independent component analysis and adjusted outliers. *International Journal of Production Research*, **48(6)**, 1727-1743.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, **10(3)**, 626-634.

Jia, M.X., Chu, F., Wang, F.L., and Wang, W. (2010). On-line batch process monitoring using batch dynamic kernel principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **101(2)**, 110–122.

Jiang, Q.C., and Yan, X.F. (2012). Chemical processes monitoring based on weighted principal component analysis and its application. Chemometrics and Intelligent Laboratory Systems, **119**, 11-20.

Lee, J.M., Qin, S.J., and Lee, I.B. (2006). Fault detection and diagnosis of multivariate process based on modified independent component analysis. *AIChE*, **52(10)**, 3501-3514.

Lee, J.M., Yoo, C.K., and Lee, I.B. (2004). Statistical process monitoring with independent component analysis. *Journal of Process Control*, **14(5)**, 467-485.

Lee, J., Kang, B., and Kang, S.H. (2011). Integrating independent component analysis and local outlier factor for plant-wide process monitoring. *Journal of Process Control*, **21(7)**, 1011-1021.

Mahadevan, S., and Shah, S.L. (2009). Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control*, **19(10)**, 1627-1639.

Odiowei, P.P., and Cao, Y. (2010). State-space independent component analysis for nonlinear dynamic process monitoring. *Chemometrics and Intelligent Laboratory Systems*, **103(1)**, 59-65.

Rashid, M.M., and Yu, J. (2012). Hidden Markov model based adaptive independent component analysis approach for complex chemical process monitoring and fault detection. *Ind. Eng. Chem. Res.*, **51(15)**, 5506-5514.

Russell, E.L., Chiang, L.H., and Braatz, R.D. (2000). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **51(1)**, 81-93.

Shao, J.D., Rong, G., and Lee, J.M. (2009). Learning a data-dependent kernel function for KPCA based nonlinear process monitoring. *Chemical Engineering Research and Design*, **87(11)**, 1471-1480.

Stefatos, G., and Hamza, A.B. (2010). Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Systems with Applications*, **37(12)**, 8606-8617.

Stubbs, S., Zhang, J., and Morris, J. (2012). Fault detection in dynamic processes using a simplified monitoring-specific CVA state space modeling approach. *Computers and Chemical Engineering*, **41(11)**, 77-87.

Tian, X.M., Zhang, X.L., Deng, X.G., and Chen, S. (2009). Multiway kernel independent component analysis based on feature samples for batch process monitoring. *Neurocomputing*, **72(7)**, 1584-1596.

Wang, J.C., Zhang, Y.B., Cao, H., and Zhu, W. Z. (2012). Dimension reduction method of independent component analysis for process monitoring based on minimum mean square error. *Journal of Process Control*, **22(2)**, 477-487.

Wang, Z.F., and Yuan, J.Q. (2007). Online supervision of Penicillin cultivations based on rolling MPCA. *Chinese Journal of Chemical Engineering*, **15(1)**, 92-96.

Ye, L., Shi, X., and Liang, J. (2011). A Multi-level approach for complex fault isolation based on structured residuals. *Chinese Journal of Chemical Engineering*, **19(3)**, 462-472.

Yu, J. (2011). Localized fisher discriminant analysis based complex chemical process monitoring. *AIChE*, **57(7)**, 1817-1828.

Yu, J.B. (2012). Local and global principal component analysis for process monitoring. *Journal of Process Control*, **22(7)**, 1358-1373.

Zhang, X., Ma, S.L., Yan, W.W., Zhao, X., and Shao, H.H. (2009). A novel systematic method of quality monitoring and prediction based on FDA and kernel regression. *Chinese Journal of Chemical Engineering*, **17(3)**, 427-436.

Zhang, Y.W., and Zhang, Y. (2010). Fault detection of non-Gaussian processes based on modified independent component analysis. *Chemical Engineering Science*, **65(16)**, 4630-4639.

Zhang, Y.W., and Ma, C. (2011). Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS. *Chemical Engineering Science*, **66(1)**, 64-72.

Zhang, Y.W., An, J.Y., and Zhang, H.L. (2013). Monitoring of time-varying processes using kernel independent component analysis. *Chemical Engineering Science*, **88**, 23-32.

Zhao, C.H., Wang, F.L., Mao, Z.Z., Lu, Y.N., and Jia, M.X. (2008). Adaptive monitoring based on independent component analysis for multiphase batch processes with limited modeling data. *Ind. Eng. Chem. Res.*, **47(9)**, 3104-3113.