Korean J. Chem. Eng., 26(6), 1467-1475 (2009)
DOI: 10.1007/s11814-009-0283-7

Two-level multiblock statistical monitoring for plant-wide processes

Zhiqiang Ge† and Zhihuan Song

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control,
Zhejiang University, Hangzhou 310027, Zhejiang, China
(*Received 25 February 2009 • accepted 18 May 2009*)

Abstract−Due to the complexity of plant-wide processes, many of the current multivariate statistical process monitoring techniques are lacking in interpretation of the detected fault, and fault identification also becomes difficult. A new two-level multiblock independent component analysis and principal component analysis (MBICA-PCA) method is proposed in this paper. Different from the conventional method, the new approach can incorporate block information into the high level for global process monitoring. Through the new method, the process monitoring task can be greatly reduced and the interpretation for the process can be made more quickly. When a fault is detected, a two-step fault identification method is proposed. The responsible sub-block is first identified by contribution plots, which is followed by fault reconstruction in the corresponding sub-block for advanced fault identification. A case study of the Tennessee Eastman (TE) process evaluates the feasibility and efficiency of the proposed method.

Key words: Plant-wide Process Monitoring, Two-level Multiblock ICA-PCA, Fault Identification, Fault Reconstruction

INTRODUCTION

As a data-driven process monitoring methodology, multivariate statistical process controls (MSPC) such as principal component analysis (PCA) and partial least squares (PLS) have been intensively researched and applied to chemical plants. There have been representative researches on process monitoring and diagnosis based on PCA and PLS for continuous [1], batch [2-4], dynamic [5], multiscale [6], and adaptive [7,8] processes. Other applications are also widely reported [9-12].

However, most of the modern chemical processes are always very complex with process variables coming from many different processing units. Those processes are known as plant-wide processes. Process monitoring and diagnosis of modern plant-wide processes become complicated and the results obtained from traditional MSPC methods are always difficult to interpret. In the last decade, hierarchical and multiblock approaches were developed, which divide the total variable block into several meaningful sub-blocks [13-18]. However, most of the multiblock methods do not consider the cross-information between divided sub-blocks. Therefore, it is possible that a deviation of the cross-information will not be observed within any sub-blocks.

Another limitation of the traditional MSPC method lies in its Gaussian distributed assumption of process variables. In fact, some of the process variables may be non-Gaussian, and chemical processes are usually driven by fewer essential variables which may not be measured. Independent component analysis (ICA) is an emerging technique for finding several independent and non-Gaussian variables as linear combinations of measured variables. A number of applications of ICA have been reported in speech processing, biomedical signal processing, machine vibration analysis, nuclear magnetic resonance spectroscopy, infrared optical source separation, radio communications and so on. [19]. Lee et al. [20,21] used ICA for process monitoring and extended it to dynamic processes. Kano et al. [22-24] developed a unified framework for MSPC, which combined PCA-based SPC and ICA-based SPC. Since the proposed combined MSPC (CMSPC) could monitor both Gaussian and non-Gaussian information of the process, good performance was shown in a multivariate system and a CSTR process. Our previous work also proposed a two-step information extraction strategy based on ICA-PCA [25].

After the fault has been detected, the fault identification step should be carried out. Conventional fault identification methods include contribution plot based methods [26], fault subspace based methods [27], and so on. However, the contribution plot based method can only narrow down the root cause of the detected fault. To determine the root cause of the fault more accurately, the reconstruction-based method can be employed [28-31]. While the reconstruction methods were developed in principal component subspace (PCS) [29] and residual subspace (RS) [27], this method has not yet been developed in the independent component subspace (ICS). By employing this method, the fault subspace should be preliminarily defined. However, while it is easy to derive the fault subspace for sensor faults, the derivation of the fault subspace is not straightforward for process faults without any knowledge. To address this problem, a subspace extraction method was proposed to extract fault directions from historical fault data [32].

In the present paper, a new plant-wide process monitoring strategy is proposed, which is based on two-level multiblock ICA-PCA. The total process variables are firstly divided into several sub-blocks, and sub-models are built correspondingly. Then a global ICA-PCA model is built on the high level for global process monitoring. The cross-information between sub-blocks can be extracted by the high level model, and also it can enhance further dimensionality reduction for process variables. Once a fault is detected, contribution plots

†To whom correspondence should be addressed.
E-mail: zqge@iipc.zju.edu.cn

are first used to calculate the responsibility of each sub-block. Then the reconstruction-based method is carried out in the responsible sub-block for advanced fault identification. The contributions of the present paper include: (1) a two-level multiblock model is proposed; (2) a new reconstruction based method in the independent component subspace is developed, and (3) a two-step fault identification method is proposed to facilitate fault identification in plant-wide processes. The rest of this paper is organized as follows. In section 2, some preliminary materials are briefly described, including the ICA, PCA and ICA-PCA methods. Section 3 demonstrates the two-level multiblock ICA-PCA method, which is followed by the fault detection, reconstruction and identification scheme in the next section. A case study of the TE benchmark process is presented in section 5, and finally some conclusions are made.

PRELIMINARIES

1. Independent Component Analysis (ICA)

ICA was originally proposed to solve the blind source separation problem [19]. To introduce the ICA algorithm, it is assumed that l measured variables, $\mathbf{x}(k)=[x_1(k), x_2(k), \ldots, x_l(k)]$ at sample k can be expressed as linear combinations of $r(\leq l)$ unknown independent components $[s_1, s_2, \ldots, s_r]^T$, and the relationship between them is given by

$$\mathbf{X}=\mathbf{A}\cdot\mathbf{S}+\mathbf{E} \tag{1}$$

where n is the number of measurements, $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]\in \mathbf{R}^{l\times n}$ is the data matrix, $\mathbf{A}=[\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_r]\in \mathbf{R}^{l\times r}$ is the mixing matrix, $\mathbf{S}=[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_r]\in \mathbf{R}^{r\times n}$ is the independent component matrix, $\mathbf{E}\in \mathbf{R}^{l\times n}$ and is the residual matrix. The basic problem of ICA is to estimate the original component $\mathbf{S}$ and the mixing matrix $\mathbf{A}$ from $\mathbf{X}$. Hyvärinen [33] introduced a very simple and efficient fixed-point algorithm (fastica) for ICA calculation.

2. Principal Component Analysis (PCA)

PCA is one of the popular multivariate statistical methods for process monitoring. Its principle is to find combinations of variables that capture the largest amount of information in the dataset. The eigenvalues of the covariance matrix are arranged in descending order--if the first k principal components (PCs) are selected, the PCA model is built on these PCs--and two statistics (T^2 and SPE) are built for monitoring. In the present paper, PCA is carried out upon the Gaussian information matrix after the non-Gaussian information has been extracted by ICA. Thus $\mathbf{E}$ can be decomposed as follows:

$$\mathbf{E}=\mathbf{T}\cdot\mathbf{P}^T+\mathbf{F} \tag{2}$$

where $\mathbf{T}$ is score matrix, $\mathbf{P}$ is loading matrix, and $\mathbf{F}$ is the residual matrix after the analysis of PCA.

3. Two-step Information Extraction Strategy

Most of the process variables contain not only Gaussian information, but also non-Gaussian information. ICA is efficient to extract essential variables, which are non-Gaussian and independent of each other. After the non-Gaussian information is extracted from the process, the rest of the Gaussian part should also be analyzed. Combining these two steps together, the original dataset $\mathbf{X}$ can be recalculated as:

$$\mathbf{X}=\mathbf{A}\cdot\hat{\mathbf{S}}+\mathbf{T}\cdot\mathbf{P}^T+\mathbf{F} \tag{3}$$

MULTIBLOCK ICA-PCA (MBICA-PCA)

Assume the data matrix $\mathbf{X}\in \mathbf{R}^{l\times n}$, where n is the number of observations and l is number of variables. According to the plant-wide process, the number of variables l is always very large. Although conventional MSPC methods show good efficiency for a multi-correlated dataset, they are limited in fault diagnosis and identification. Results obtained from these conventional MSPC methods are often difficult to interpret. To improve fault detection and identification for large complex processes, a new two-level multiblock method is proposed. Process variables are divided into several sub-blocks based on prior knowledge:

$$\mathbf{X}=[\mathbf{X}_1\ \ \mathbf{X}_2\ \ \cdots\ \mathbf{X}_B] \tag{4}$$

where B is the number of sub-blocks, each sub-block $\mathbf{X}_b\in \mathbf{R}^{n\times m_b}$ (b= 1, 2, $\cdots$, B) has m_b variables. The main idea of the two-level multiblock ICA-PCA method is described as follows. A sub-model is built for each of the sub-block; then a global model is built on a high level, in which non-Gaussian and Gaussian information are modeled separately. The proposed two-level multiblock ICA-PCA method is illustrated in Fig. 1.

In the conventional multiblock PCA algorithm [15], each variable in the data blocks $\mathbf{X}_b$ is often scaled to have zero mean and variance of $1/m_b$ to make each block contribute about the same variance to the super scores. For simplicity, we assume the dataset $\mathbf{X}$ has already been scaled. Therefore, B sub-models are first built for corresponding sub-blocks $\mathbf{X}_b$ (b=1, 2, $\cdots$, B).

$$\mathbf{X}_b=\mathbf{A}_b\cdot\mathbf{S}_b+\mathbf{E}_b \tag{5}$$

$$\mathbf{E}_b=\mathbf{T}_b\cdot\mathbf{P}_b^T+\mathbf{F}_b \tag{6}$$

It is important to note that some sub-blocks only contain the Gaussian information, which can be judged by negentropy [19]. For these sub-blocks, PCA is enough for modeling the correlation within these sub-blocks. After all of the B sub-models have been built, the extracted independent components and principal components in each sub-block are arranged as follows.

$$\mathbf{S}_{mix}=[\mathbf{S}_1\ \ \mathbf{S}_2\ \ \cdots\ \ \mathbf{S}_B]^T \tag{7}$$

$$\mathbf{T}_{mix}=[\mathbf{T}_1\ \ \mathbf{T}_2\ \ \cdots\ \ \mathbf{T}_B] \tag{8}$$

Although the correlations within each sub-block are well extracted, the correlations between sub-blocks are not well extracted. Therefore, a global model could be built on a high level to extract the cross-

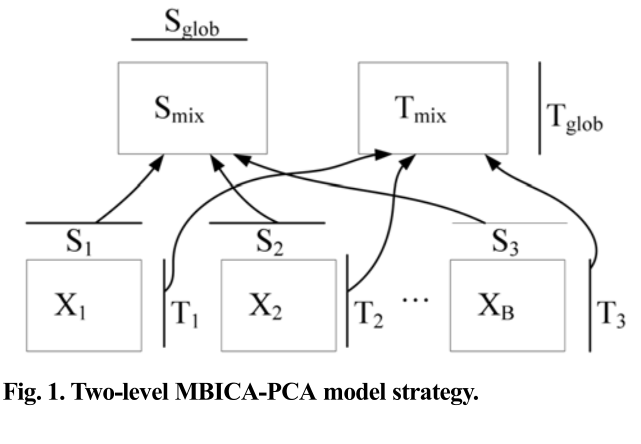

Fig. 1. Two-level MBICA-PCA model strategy.

information, and also to model non-Gaussian and Gaussian information separately. According to the newly defined data matrix $\mathbf{S}_{mix}$ and $\mathbf{T}_{mix}$, the global ICA-PCA model is built as follows.

$$\mathbf{S}_{mix}=\mathbf{A}_{glob}\cdot\mathbf{S}_{glob}+\mathbf{E}_{glob} \tag{9}$$

$$\mathbf{T}_{mix}=\mathbf{T}_{glob}\cdot\mathbf{P}_{glob}^{T}+\mathbf{F}_{glob} \tag{10}$$

Note that block dividing plays an important role in the multiblock method. It is therefore important to make intelligent selections of variable blocking based on process knowledge such that the most important correlations are extracted within divided sub-blocks. When the built model is used for process monitoring and fault diagnosis, it is much easier to isolate and identify the detected fault. Furthermore, with the development of the global model, cross-information between sub-blocks can be efficiently extracted. The variable dimension may also be reduced at the second time. Hence, more redundant information can be removed from the monitoring information, which will also enhance the performance of process monitoring.

PROCESS MONITORING BASED ON TWO-LEVEL MBICA-PCA

In this section, the proposed method is demonstrated, including fault detection and identification. Suppose we have already built all of the models described in section 3, including a global model and B sub-models. To monitor the Gaussian and non-Gaussian information separately, three statistics will be established for each ICA-PCA model. If the statistical confidence limits are exceeded, some disturbance or fault may happen in the process. Then the contribution plot of each statistic can be used to identify the sub-block responsible for the abnormal event. Then, advanced fault identification can be carried out in the responsible sub-block to find the root cause of the detected fault.

1. Fault Detection

Given the sub-block data matrix $\mathbf{X}_b$, suppose r_b independent components are extracted, $\mathbf{S}_b=[\mathbf{s}_{b1}, \mathbf{s}_{b2}, \cdots, \mathbf{s}_{bn}]\in\mathbf{R}^{r_b\times n}$. To monitor the non-Gaussian part of the process, the I^2 statistic variable is defined [20]:

$$I_b^2=\mathbf{s}_b^T\cdot\mathbf{s}_b \tag{11}$$

After the non-Gaussian information has been extracted, the residual matrix $\mathbf{E}_b$ is obtained. As mentioned in section 3, we use PCA to analyze it, expanding $\mathbf{E}_b$ as below:

$$\mathbf{E}_b=\sum_{i=1}^{k_b}\mathbf{t}_i\cdot\mathbf{p}_i^T+\mathbf{F}_b \tag{12}$$

where k_b is the number of principal components, $\mathbf{F}_b$ is the residual resulting from the PCA model. Here we define the limits of T_b^2 and SPE_b statistics as follows [1,2]:

$$T_b^2=\sum_{i=1}^{k_b}\frac{\mathbf{t}_i\mathbf{t}_i^T}{\lambda_i}\le\frac{k_b(n-1)}{n-k_b}F_{k,(n-k_b),\alpha} \tag{13}$$

$$SPE_b=\mathbf{f}\cdot\mathbf{f}^T=\mathbf{e}(\mathbf{I}-\mathbf{P}_b\mathbf{P}_b^T)\mathbf{e}^T\le SPE_\alpha \tag{14}$$

$$SPE_\alpha=\theta_1\cdot\left[1+\frac{c_\alpha\sqrt{2\theta_2h_0^2}}{\theta_1}+\frac{\theta_2h_0(h_0-1)}{\theta_1^2}\right]^{1/h_0} \tag{15}$$

where k_b is the number of PCs, $\theta_i=\sum_{j=k+1}^{m}\lambda_j^i$ for i=1, 2, 3, $h_0=1-(2\theta_1\theta_3/3\theta_2^2)$, α is significance level, c_α is the normal deviate corresponding to the upper $1-\alpha$ percentile.

On the high level of the model, the monitoring data matrices become the new arranged data matrices $\mathbf{S}_{mix}$ and $\mathbf{T}_{mix}$; similar statistics can be established. Suppose $\mathbf{S}_{glob}$ is the independent component matrix calculated by the global ICA model, and $\mathbf{T}_{glob}$ is the score matrix of $\mathbf{T}_{mix}$ which is calculated by the global PCA model. The three statistics are established as below.

$$I_{glob}^2=\mathbf{S}_{glob}^T\cdot\mathbf{S}_{glob} \tag{16}$$

$$T_{glob}^2=\mathbf{t}_{glob}\cdot\Lambda_{glob}\cdot\mathbf{t}_{glob}^T \tag{17}$$

$$SPE_{glob}=\mathbf{f}_{glob}\cdot\mathbf{f}_{glob}^T \tag{18}$$

In PCA monitoring, the confidence limits are based on a specified distribution shown in Eqs. (13)-(15) based upon the assumption that the latent variables follow a Gaussian distribution. However, in ICA monitoring, the independent component does not conform to a specific distribution. Hence, the confidence limit of the I^2 statistic cannot be determined directly from a particular approximate distribution. An alternative approach to define the nominal operating region of the I^2 statistic is to use kernel density estimation (KDE) [34,35]. For simplicity, the confidence limit of I^2 can also be determined by trial and error.

2. Fault Identification

After a fault has been detected, the contribution plot method is first selected to identify the responsible sub-block. For advanced fault identification, the reconstruction-based method is carried out in each of the three subspaces: independent component subspace (ICS), principal component subspace (PCS), and residual subspace (RS). Therefore, the root cause of the detected fault can be further identified, and the fault interpretation is improved. To determine which block takes the responsibility for the fault, block contributions are defined as follows, which are average values of the variable contributions in corresponding blocks.

$$I^2Cont_{block,b}=\sum_{i\in VI(b)}I^2Cont_{glob,i}/r_b(b=1,2,\cdots,B) \tag{19}$$

$$T^2Cont_{block,b}=\sum_{i\in VP(b)}T^2Cont_{glob,i}/k_b(b=1,2,\cdots,B) \tag{20}$$

$$SPECont_{block,b}=\sum_{i\in VP(b)}SPECont_{glob,i}/k_b(b=1,2,\cdots,B) \tag{21}$$

where r_b is the number of independent components extracted from block b, k_b is the number of principal component extracted from block b. VI(b) means the independent component number extracted in the b-th sub-block, VP(b) and means the principal component number extracted in the b-th sub-block. The contributions of the i-th component in Eqs. (19)-(21) are defined as [20,26]:

$$I^2Cont_{glob,i}=\frac{\mathbf{A}_{glob}\cdot\mathbf{S}_{fault,glob,i}}{\|\mathbf{A}_{glob}\cdot\mathbf{S}_{fault,glob,i}\|}\|\mathbf{S}_{fault,glob,i}\| \tag{22}$$

$$T^2Cont_{glob,i}=\sum_{i=1}^{k}\frac{\mathbf{t}_i}{\lambda_i}\mathbf{P}_{glob,i}\cdot\mathbf{e}_{fault,glob} \tag{23}$$

$$SPECont_{glob,i}=(\mathbf{f}_{fault,glob,i}-\hat{\mathbf{f}}_{fault,glob,i})^2 \tag{24}$$

In the present paper, the fault subspace is supposed to be known.

Otherwise, it can be extracted by the singular vector decomposition (SVD) method proposed by Yue et al. [32]. Suppose the fault set includes J faults, which is described as $\{F_j, J=1, 2, \cdots J\}$. Denote $\{\Theta_j, j=1, 2, \cdots J\}$ as the fault subspace with dimensions $\{\dim(\Theta_j)\geq 1, j=1, 2, \cdots J)$ for the defined fault set. Hence, both of the unidimensional and multidimensional faults are considered.

After the fault subspace has been defined, the normal value $\mathbf{x}^*$ can be reconstructed from the corrupted value $\mathbf{x}$ in each of the IC, PC, and residual subspaces. Suppose a fault F_j has happened; a reconstructed value $\mathbf{x}_j$ can be obtained as an adjustment of the corrupted value $\mathbf{x}$ moving along a given fault direction Θ_j:

$$\mathbf{x}_j=\mathbf{x}-\Theta_j\mathbf{f}_j \tag{25}$$

where $\mathbf{f}_j$ is the estimated fault vector with $\|\mathbf{f}\|$ as its magnitude such that $\mathbf{x}_j$ is closest to the normal region.

In the IC subspace, theoretically, the optimal reconstruction is obtained by minimizing $\|\mathbf{x}_j-\mathbf{x}^*\|$. However, this is infeasible because the normal value $\mathbf{x}^*$ is unknown. In the present paper, the reconstruction is realized by moving the corrupted value $\mathbf{x}$ along the defined fault direction Θ_j in ICS. The reconstruction can be formulized in the following optimization problem:

$$\mathbf{f}_j=\arg\lim_{f}\|\mathbf{W}(\mathbf{x}-\Theta_j\mathbf{f})\|^2 \tag{26}$$

Denoting $\Delta_j=\mathbf{W}\Theta_j$, the solution of Eq. (26) is straightforward by least square technique, which yields the following:

$$\mathbf{f}_j=(\Delta_j^T\Delta_j)^{-1}\Delta_j^T\mathbf{W}\mathbf{x} \tag{27}$$

where the column rank of the matrix Δ_j is assumed to be full. If the matrix Δ_j suffers a column rank deficiency, then the corresponding solution of the optimization problem in Eq. (26) will be

$$\mathbf{f}_j=\Delta_j^+\mathbf{W}\mathbf{x} \tag{28}$$

where the matrix Δ_j^+ is the Moore-Penrose pseudo-inverse of Δ_j. Hence, the reconstructed data vector can be represented as Eq. (25). The new I^2 statistic value of this reconstructed data sample can be calculated as follows:

$$I_j^2=\mathbf{S}_j^T\mathbf{S}_j \tag{29}$$

where $\mathbf{S}_j=\mathbf{W}\mathbf{x}_j$, then the identification index in IC subspace is defined as

$$\eta_{ICS,j}=\frac{I_j^2}{I_{fault}^2} \tag{30}$$

where I_{fault}^2 is the statistical value with faulty data. Therefore, when the fault subspace Θ_j is matched, the value of $\eta_{ICS,j}$ will be reduced significantly.

Similarly, the detected fault can be reconstructed in both PCS and RS. Related works have been published, including Qin [27], Wang [29], and Lieftuche [30,31]. Therefore, the T^2 and SPE statistic values of the reconstructed data sample can be calculated, respectively, in PCS and RS:

$$T_j^2=\|\lambda^{-1/2}\mathbf{P}^T\mathbf{x}_j\|^2 \tag{31}$$

$$SPE_j=\|(\mathbf{I}-\mathbf{P}\mathbf{P}^T)\mathbf{x}_j\|^2 \tag{32}$$

where λ is the eigenvalue matrix, $\mathbf{I}$ is an identity matrix with appropriate dimension. Then the identification index in PC and residual subspaces can be defined as

$$\eta_{PCS,j}=\frac{T_j^2}{T_{fault}^2} \tag{33}$$

$$\eta_{RS,j}=\frac{SPE_j}{SPE_{fault}} \tag{34}$$

In summary, two steps are taken for fault identification. First, the contribution plot method is used to identify the responsible sub-block for the detected fault. Although with this method it is hard to pinpoint the specific cause of the fault, it can narrow down the possible causes to a specific sub-block or several sub-blocks. After the responsible sub-block has been determined, fault reconstruction and identification can be carried out in any of the three subspaces (ICS, PCS, and RS) as long as it can be detected by its corresponding statistic.

CASE STUDY

In this section, the proposed method is tested through the TE process [36]. As a benchmark simulation, the TE process has been widely used to test the performance of various monitoring approaches [37]. This process has 41 measured variables (22 continuous process measurements and 19 composition measurements) and 12 manipulated variables, and a set of 21 programmed faults are introduced to the process. The details on the process description are well explained in by Chiang et al. [37]. In the present paper, 33 variables are selected for process monitoring, which are listed in Lee et al. [21]. The simulation data which we have collected are separated into two parts: training datasets and test datasets. They both consisted of 960 observations for each operation mode, and their sampling interval was 3 min. All process faults are introduced in the process after sample 160.

Previously, it has been logical to divide the process into sub-blocks that describe a unit or a specific physical or chemical operation. Since the TE process consists of five major units, we can divide the process variables into five sub-blocks, each block corresponding to one unit. However, because the condenser unit and the compressor unit only have two variables, they are suggested to be integrated into the other three sub-blocks. Therefore, these 33 monitoring variables are divided into three sub-blocks. There are 16 (1-9, 21, 23-26, 32, 33), 10 (10-14, 20, 22, 27-29) and 7 (15-19, 30, 31) variables in these three sub-blocks. To build the sub-model in the first sub-block, 4 independent components (ICs) and 6 principal components (PCs) are selected. The other two selections of the corresponding sub-models are 3 ICs, 4 PCs and 2 ICs, 3 PCs. On the high level modeling, 7 independent components and 10 principal components are selected. To examine the rationality of our choices of component numbers, the Jarque-Bera test is introduced. In this paper, the *jbtest* function in MATLAB toolbox is used. If the value of the *jbtest* func-

Table 1. Jarque-Bera test results of independent and principal components

Components	IC_{glob4}	IC_{glob5}	IC_{glob6}	IC_{glob7}	PC_{glob1}	PC_{glob2}	PC_{glob3}	PC_{glob4}
H	1	1	1	1	0	0	0	0
Components	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8
H	1	1	1	1	1	1	1	1

tion equals to 1, it means that the tested component is non-Gaussian. On the other hand, if the value is zero, the tested component is considered to be Gaussian. Results of the Jarque-Bera test are tabulated in Table 1. For comparison, principal components of conventional PCA are also tested. Results of the last 4 ICs and the first 4 PCs of the global model are shown in the second row of Table 1, and results of the first 8 PCs of the PCA model are given in the last row of the table. As shown in the table, the values of IC_{glob7} and PC_{glob1} are 1 and 0, which means that the non-Gaussian information is approximately extracted by 7 ICs. However, the first 8 PCs of PCA are all considered non-Gaussian. In fact, if 19 PCs are chosen for the PCA model, among which the first 12 PCs are considered to be non-Gaussian, only 7 PCs (13-19) are Gaussian distributed.

To evaluate the monitoring performance of the proposed method, a sensor fault is first simulated in the TE process. A constant value is added to variable 5, and 960 samples are collected. Global process

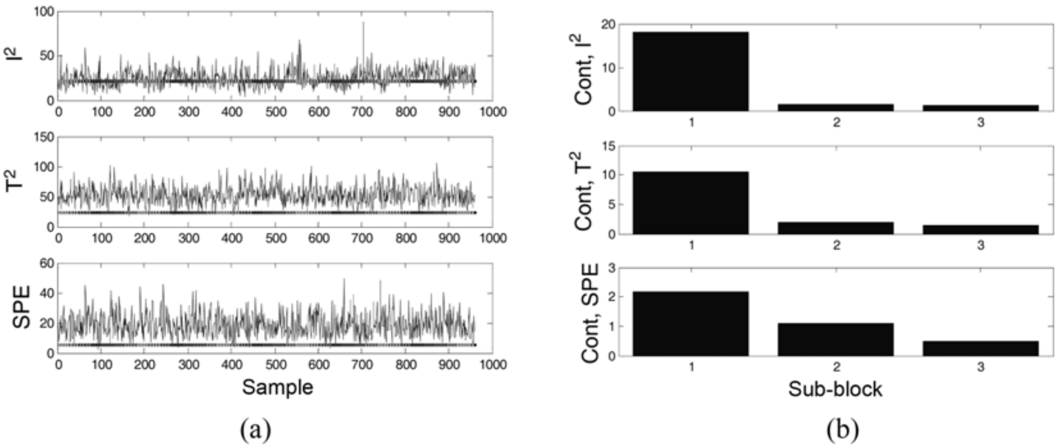

Fig. 2. Monitoring results of the sensor fault: (a) fault detection; (b) sub-block identification.

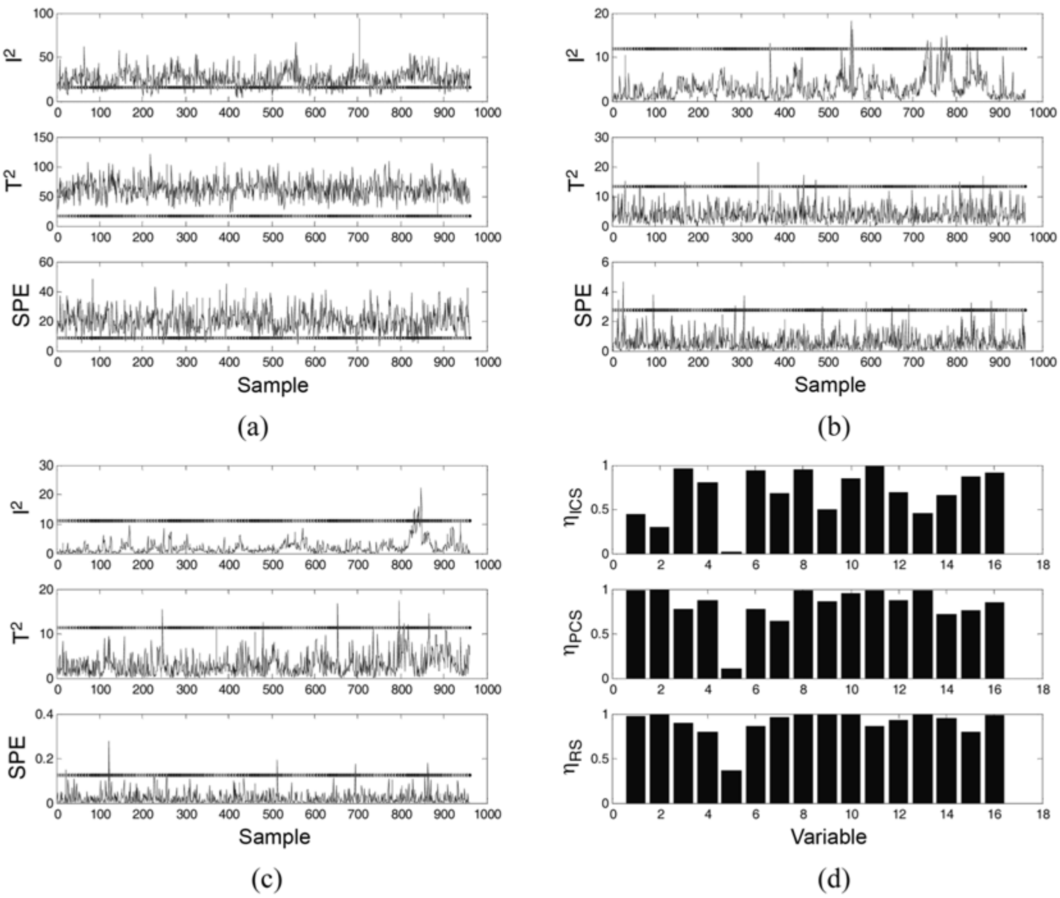

Fig. 3. Advanced monitoring results of the sensor fault: (a) fault detection results of sub-block 1; (b) fault detection results of sub-block 2; (c) fault detection results of sub-block 3; (d) advanced fault identification results of the sensor fault.

monitoring results of this sensor fault are given in Fig. 2. As indicated, all three statistical confidence limits have been exceeded, which means that a fault has been detected. To determine which sub-block takes the most responsibility for the fault, the contribution plot method is employed. It is clearly shown in Fig. 2(b) that the first sub-block should be the most responsible one for the detected fault. The results of contribution plots are straightforward and easy to interpret, since the faulty sensor is included in the first sub-block. Monitoring results of three sub-blocks are given in Fig. 3. It can be inferred that this sensor fault has happened in the first sub-block, since most of the statistical values exceed their corresponding control limits. However, the statistical confidence limits in Fig. 3(b) and Fig. 3(c) are not violated during the operating process. To gain further interpretation and determine the exact root cause of the fault, advanced fault detection and identification can be carried out. Thus the reconstruction-based method is employed in the first sub-block. Final

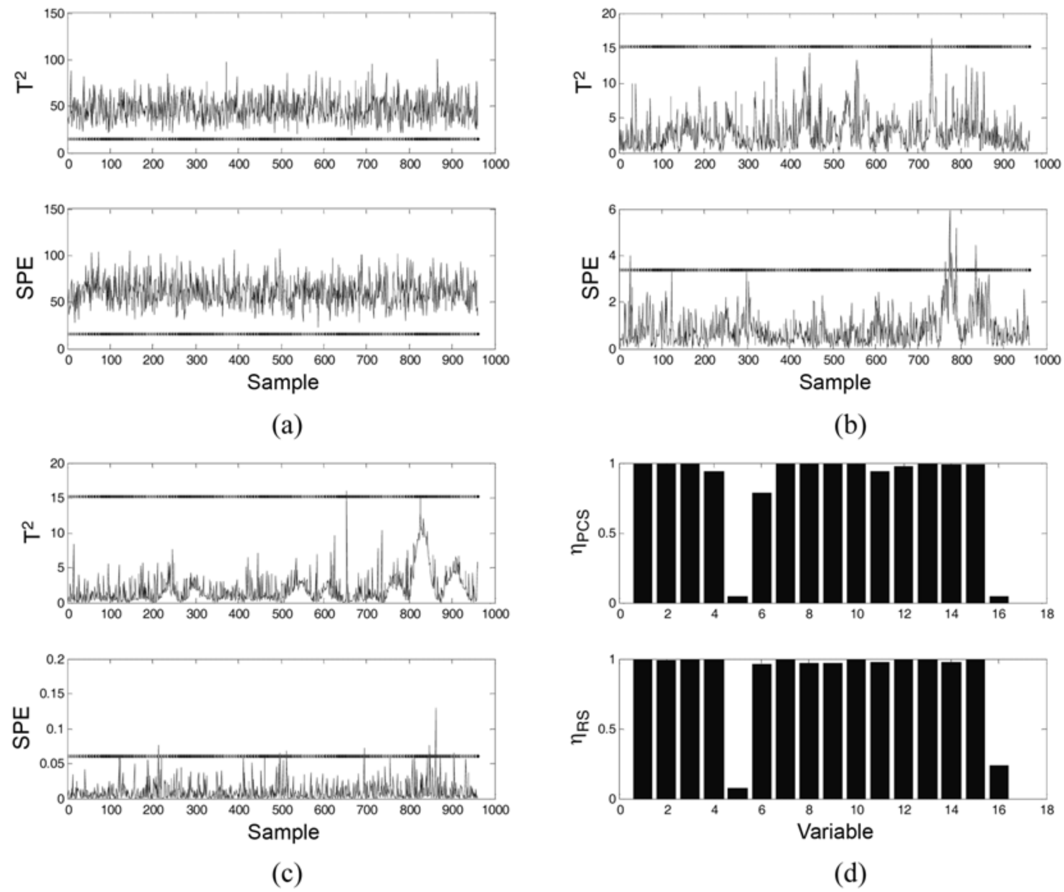

Fig. 4. Monitoring results of the sensor fault by Multiblock PCA: (a) fault detection results of sub-block 1; (b) fault detection results of sub-block 2; (c) fault detection results of sub-block 3; (d) fault identification results of the sensor fault.

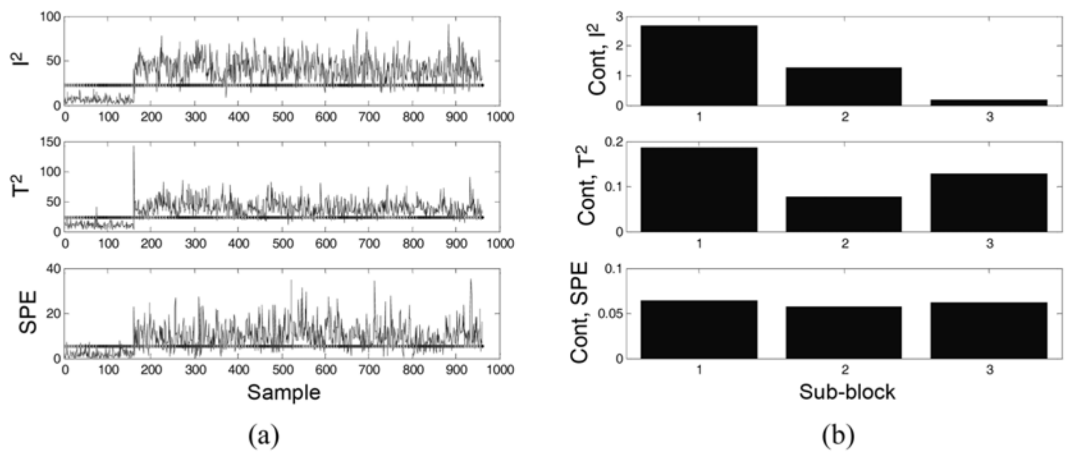

Fig. 5. Monitoring results of fault 4: (a) fault detection; (b) sub-block identification.

identification results are shown in Fig. 3(d). One can find that the faulty variable (variable 5) is correctly identified in any of the three subspaces (ICS, PCS and RS). For comparison, monitoring results of multiblock PCA are given in Fig. 4. Because no super model has been built for global process monitoring, three sub-blocks should be monitored separately. If there are many sub-blocks, the monitoring tasks will become troublesome. However, if the two-level multiblock model is used, one can first use the global model to monitor the whole process, and then the advanced process monitoring should only be done in the responsible sub-block. Although the fault detection and identification results of the two methods seem to be identical in this case, the two-level multiblock method has more potential for monitoring plant-wide processes.

Next, fault 4 is considered. Unlike the sensor fault, whose fault subspace is easily derived, fault subspace derivations of process faults are not straightforward. To this end, Yue et al. [32] proposed a fault subspace extraction strategy for the reconstruction-based fault identification method. Before carrying out fault identification, all of the 21 fault subspaces in the TE process are supposed to be extracted and preserved. Fault detection and identification results of both methods for fault 4 are shown in Figs. 6-8. As seen in Fig. 5, three monitoring statistics all indicate that a fault has happened, and the contribution plots given in Fig. 5(b) indicate that the first sub-block takes the most responsibility. Therefore, the first sub-block is further monitored and advanced fault identification is also carried out in this sub-block. To illustrate the responsibility of the first sub-block, fault detection results of the second and third sub-blocks are also given in Fig. 6. Besides, the results of advanced fault identification are also given in Fig. 6(d), in which the fourth fault subspace is found to be the most likely one that could happen in the process. However, the 11-th fault subspace seems to be similar to the identified fault subspace, which makes the fault identification result puzzling. In this case, the performance could be improved by incorporating some process or expert knowledge. Similar fault detection results are obtained by multiblock PCA, which are given in Fig. 7. However, the fault identification results seem to be worse than that in Fig. 6(d).

CONCLUSIONS

A novel strategy has been developed for plant-wide process monitoring. The new proposed method is based on two-level multiblock ICA-PCA. The multiblock method first divides the process into several sub-blocks. Sub-models are built for these divided sub-blocks. Then the extracted information from these sub-blocks is integrated on the high level. A global monitoring model is developed on this high level to extract further cross-information between sub-blocks. When a fault has been detected by the global monitoring charts, a two-step fault identification method is proposed, which is based on the contribution plots and the fault reconstruction method. Compared to the conventional approach, the new proposed method is

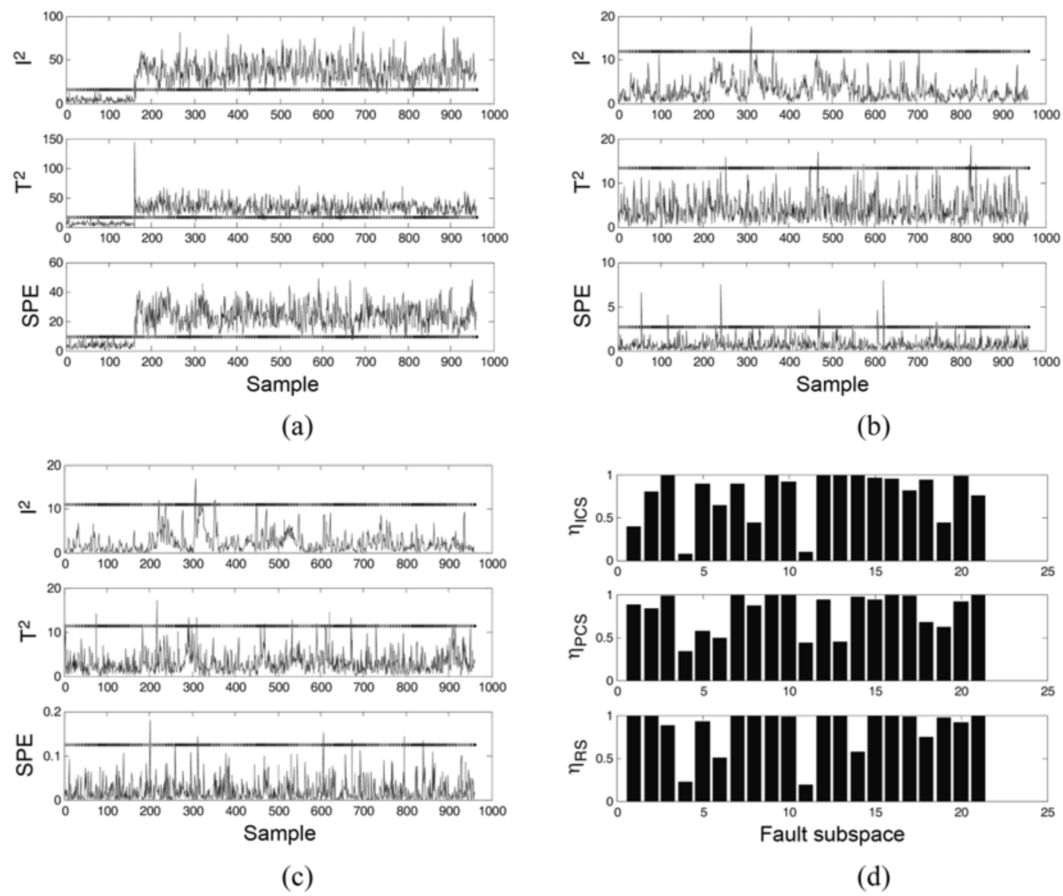

Fig. 6. Advanced monitoring results of fault 4: (a) fault detection results of sub-block 1; (b) fault detection results of sub-block 2; (c) fault detection results of sub-block 3; (d) advanced fault identification results of fault 4.

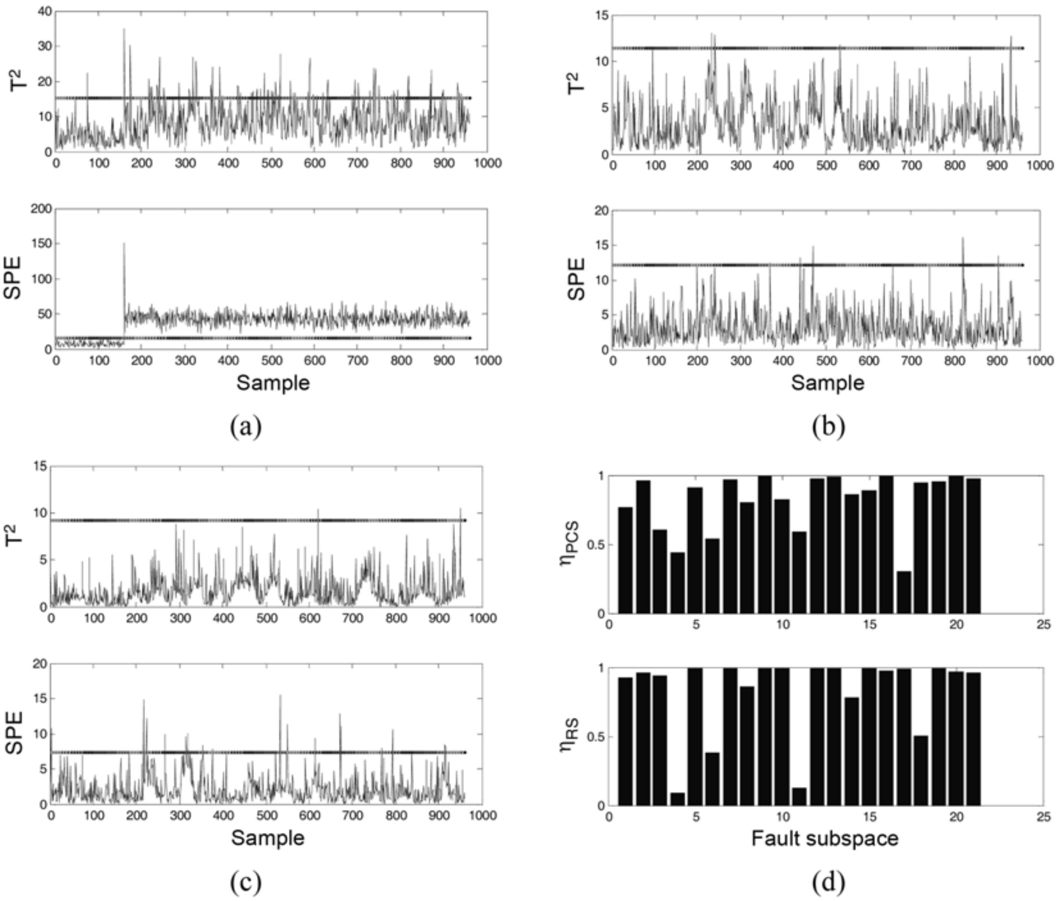

Fig. 7. Monitoring results of fault 4 by Multiblock PCA: (a) fault detection results of sub-block 1; (b) fault detection results of sub-block 2; (c) fault detection results of sub-block 3; (d) fault identification results of fault 4.

more efficient for both fault detection and identification. The TE case study shows the feasibility and efficiency of the proposed method.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China. (Grant#60774067).

REFERENCES

1. J. Kresta, J. F. MacGregor and T. Marlin, *Can. J. Chem. Eng.*, **69**, 35 (1991).
2. P. Nomikos and J. F. MacGregor, *AIChE J.*, **40**, 1361 (1994).
3. J. H. Chen and J. H. Yen, *Korean J. Chem. Eng.*, **20**, 1000 (2003).
4. M. H. Kim and C. K. Yoo, *Korean J. Chem. Eng.*, **25**, 947 (2008).
5. W. Ku and R. H. Storer, *Chem. Intell. Lab. Syst*., **30**, 179 (1995).
6. B. R. Bakshi, *AIChE J*., 44, 1596 (1998).
7. S. J. Qin, *Comput. Chem. Eng*., **22**, 503 (1998).
8. W. Li, H. Yue, S. Valle and S. J. Qin, *J. Process Control.*,**10**, 471 (2000).
9. S. Lee, S. Yeom and K. S. Lee, *Korean J. Chem. Eng.*, **21**, 575 (2004).
10. I. G. Byun, H. U. Nam, S. K. Song, I. S. Hwang, T. H. Lee and T. J. Park, *Korean J. Chem. Eng.*, **22**, 917 (2005).
11. K. S. Kim and J. W. Ko, *Korean J. Chem. Eng.*, **22**, 26 (2005).
12. C. K. Lee and I. B. Lee, *Korean J. Chem. Eng.*, **25**, 203 (2008).
13. J. F. MacGregor, C. Jaeckle, C. Kiparissides and M. Kourtoudi, *AIChE J*., **40**, 826 (1994).
14. J. A. Westerhuis, T. Kourti and J. F. MacGregor, *J. Chemometrics.*, **12**, 301 (1998).
15. S. J. Qin, S. Valle and M. J. Piovoso, *J. Chemometrics.*, **15**, 715 (2001).
16. A. K. Smilde, J. A. Westerhuis and Sijmen de Jong, *J. Chemometrics.*, **17**, 323 (2003).
17. S. W. Choi and I. B. Lee, *J. Process Control.*, **15**, 295 (2005).
18. G. A. Cherry and S. J. Qin, *IEEE Trans. Semiconductor Manufacturing*, **19**, 159 (2006).
19. A. Hyvarinen and E. Oja, *Neural Network*, **13**, 411 (2000).
20. J. M. Lee, C. K. Yoo and I. B. Lee, *J. Process Control*., **14**, 467 (2004).
21. J. M. Lee, C. K. Yoo and I. B. Lee, *Chem. Eng. Sci.*, **59**, 2995 (2004).
22. M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto and H. Ohno, *Combined multivariate statistical process control*, IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM), 303 (2004).
23. M. Kano, S. Tanaka, S. Hasebe, I. Hashimoto and H. Ohno, *AIChE J.*, **49**, 969 (2003).
24. M. Kano, S. Hasebe, I. Hashimoto and H. Ohno, *Comput. Chem. Eng*., **28**, 1157 (2004).
25. Z. Q. Ge and Z. H. Song, *Ind. Eng. Chem. Res*., **46**, 2054 (2007).
26. J. Westerhuis, S. Gurden and A. Smilde, *Chem. Intell. Lab. Syst*., **51**, 95 (2000).

27. R. Dunia and S. J. Qinm, *AIChE J.*, **44**, 1813 (1998).
28. R. Dunia and S. J. Qin, *Comput. Chem. Eng.*, **22**, 927 (1998).
29. H. Q. Wang, Z. H. Song and P. Li, *Ind. Eng. Chem. Res.*, **41**, 2455 (2002).
30. D. Lieftuche, U. Kruger, G. W. Irwin and R. J. Treasure, *IEE Proc.-Control Theory Appl.*, **153**(4), 437 (2006).
31. D. Lieftuche, U. Kruger and G. W. Irwin, *Comput. Chem. Eng.*, **30**, 901 (2006).
32. H. H. Yue and S. J. Qin, *Ind. Eng. Chem. Res.*, **40**, 4403 (2001).
33. A. Hyvarinen, *Neural Computation.*, **9**, 1483 (1997).
34. Q. Chen, R. J. Wynne, P. Goulding and D. Sandoz, *Control Eng. Pract.*, **8**, 531 (2000).
35. Q. Chen, U. A. Kruger, T. Y. Leung, *Control Eng. Pract.*, **12**, 267 (2004).
36. J. J. Downs and E. F. Vogel, *Comput. Chem. Eng.*, **17**, 245 (1993).
37. L. H. Chiang, E. L. Russell and R. D. Braatz, *Fault detection and diagnosis in industrial systems*, Springle-Verlag, London (2001).