Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Batch process monitoring based on multilevel ICA-PCA*

Zhi-qiang GE, Zhi-huan SONG‡

(*State Key Lab of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, China*)

E-mail: zqge@iipc.zju.edu.cn; zhsong@iipc.zju.edu.cn

Received Nov. 6, 2007; revision accepted Mar. 10, 2008

Abstract: In this paper, we describe a new batch process monitoring method based on multilevel independent component analysis and principal component analysis (MLICA-PCA). Unlike the conventional multi-way principal component analysis (MPCA) method, MLICA-PCA provides a separated interpretation for multilevel batch process data. Batch process data are partitioned into two levels: the within-batch level and the between-batch level. In each level, the Gaussian and non-Gaussian components of process information can be separately extracted. I^2, T^2 and *SPE* statistics are individually built and monitored. The new method facilitates fault diagnosis. Since the two variation levels are decomposed, the variables responsible for faults in each level can be identified and interpreted more easily. A case study of the Dupont benchmark process showed that the proposed method was more efficient and interpretable in fault detection and diagnosis, compared to the alternative batch process monitoring method.

Key words: Multilevel, Independent component analysis (ICA), Principal component analysis (PCA), Batch process monitoring, Non-Gaussian

doi:10.1631/jzus.A0720051 **Document code:** A **CLC number:** TP273

INTRODUCTION

To meet market requirements for flexibility, batch processes have received increasing attention in recent years. Batch process monitoring approaches based on multi-way principal component analysis (MPCA) and multi-way partial least squares (MPLS) have been developed in the last decade (Nomikos and MacGregor, 1994; 1995a; 1995b). Many other techniques using multivariate statistical analysis have also been developed and applied to industrial batch process monitoring (Simoglou *et al.*, 2005; Camacho and Picó, 2006; Doan and Srinivasan, 2008). A characteristic of batch process data is that it is multilevel and can be partitioned into two levels: the within-batch level and the between-batch level. If PCA/PLS is used for the analysis of multilevel data, the different types of variation in the multilevel data will not be separated and the principal components obtained will describe a mixture of different types of variation. In the analysis of batch process data, this means that a PCA/PLS model does not give a separate interpretation of the between-batch variation and the within-batch variation of the process. Moreover, MPCA/MPLS is based on the assumption that the process variables are Gaussian and future values must be estimated to the end of the batch at each sample time.

Multilevel data are generated in research in biology, physics and many other fields. Examples of multilevel problems include the monitoring of hospital patients in time, multilevel analysis of chemical process data and economic time-series analysis of multiple countries, economic branches or companies (de Noord and Theobald, 2005; Jansen *et al.*, 2005). Kiers and Ten Berge (1994) proposed a generalization of PCA to analyze data containing sets of samples belonging to multiple populations in which the same variables were measured. This method, simultaneous component analysis (SCA), removes the static variation between the populations. SCA is focused on the

‡ Corresponding author
* Project (No. 60774067) supported by the National Natural Science Foundation of China

variation within the populations, and this variation is not confounded with the variation between populations. Timmerman and Kiers (2003) have applied this method to the analysis of multivariate, multi-subject, time-resolved data. However, because the static variation between individuals was lost from the analysis, potentially valuable information about the process underlying the variation in the data was not obtained using the SCA method. Jansen *et al.*(2005) proposed a multilevel method called 'multilevel component analysis' (MCA), which makes a model of a dataset that contains different sub-models for the different types of variation in the data. Thus, a two-level MCA model of time-resolved measurements on multiple individuals contains two sub-models: one describing the dynamic variation of the individuals and the other describing the static differences between the individuals. Many applications of these technologies have been published. However, so far there are no publications in the area of batch process monitoring.

In this paper, a novel on-line batch process monitoring strategy is proposed, called 'multilevel independent component analysis and principal component analysis' (MLICA-PCA). This new method gives a separated interpretation for the multilevel batch process data. It also extracts the non-Gaussian information from the process and monitors the non-Gaussian and Gaussian information separately. The estimation of future values is avoided. In Section 2 we discuss some preliminaries including ICA-PCA and multilevel statistical analyses, followed by the proposed MLICA-PCA method in Section 3. The on-line batch process monitoring strategy based on MLICA-PCA is described in Section 4. In Section 5, the proposed method is applied to the Dupont batch process. Finally, we summarize our conclusions.

PRELIMINARIES

Two-step information extraction strategy based on ICA-PCA

ICA is efficient for extracting essential variables which are non-Gaussian and independent of each other. It is an emerging technology used in many areas (Hyvarinen and Oja, 2000) but has been applied only recently in process monitoring (Kano *et al.*, 2003; Yoo *et al.*, 2004; Lee *et al.*, 2006). To analyze the Gaussian data after the non-Gaussian information has been extracted from the process, a two-step information extraction strategy was proposed (Ge and Song, 2007). Since the conventional PCA method can adequately model the Gaussian information of the process, ICA is carried out first to extract the non-Gaussian information and then PCA is applied. A given process dataset $\boldsymbol{X}$ can be decomposed as follows:

$$\boldsymbol{X} = \boldsymbol{A}\hat{\boldsymbol{S}} + \boldsymbol{E}, \quad (1)$$

where $\hat{\boldsymbol{S}}$ is the independent component matrix extracted by the ICA procedure in the first step, and $\boldsymbol{E}$ is the residual matrix whose Gaussian information is modeled by PCA (Nomikos and MacGregor, 1994) in the second step. $\boldsymbol{E}$ is decomposed as

$$\boldsymbol{E} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{F}, \quad (2)$$

where $\boldsymbol{T}$ is the score matrix, $\boldsymbol{P}$ is the loading matrix, and $\boldsymbol{F}$ is the residual matrix after PCA analysis. Combining the two steps together, the original dataset $\boldsymbol{X}$ can be recalculated as

$$\boldsymbol{X} = \boldsymbol{A}\hat{\boldsymbol{S}} + \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{F}. \quad (3)$$

Multilevel statistical analysis

The standardized data collected from batch processes are organized into a 3D array $I\times J\times K$, where I, J and K are the number of batches, monitoring variables and time samples over the duration of the batch, respectively. As shown in Fig.1, the original dataset is defined as $\boldsymbol{X}_{N\times J}$, where $N=\sum_{i=1}^{I}K_i$. In special cases, the batch run length K_i can be the same for every batch, but in practice they are usually different from each other. An element x_{ijk} of matrix $\boldsymbol{X}$, which contains a measurement of batch i on variable j at time k, can be decomposed as

$$x_{ijk} = x_{\cdot j\cdot} + (x_{ij\cdot} - x_{\cdot j\cdot}) + (x_{ijk} - x_{ij\cdot}), \quad (4)$$

where $x_{\cdot j\cdot}$ is the overall mean of variable j, and $x_{ij\cdot}$ is the mean of batch i on variable j (de Noord and Theobald, 2005). In the right-hand side of Eq.(4), the first term is an offset that is constant across batches and time points, the second part is a between-batch deviation, and the third part describes the within-batch deviation. In the same way, the sum of variable squares can be separated into three parts as follows:

$$\sum_{k=1}^{K}\sum_{i=1}^{I}x_{ijk}^2 = IKx_{.j.}^2 + \sum_{i=1}^{I}K(x_{ij.} - x_{.j.})^2 + \sum_{k=1}^{K}\sum_{i=1}^{I}(x_{ijk} - x_{ij.})^2. \quad (5)$$

Multilevel statistical analysis separates the original dataset into three different parts. Therefore, variation in different parts can be modeled separately and the changes occurring in each part can be analyzed. Compared to the conventional method, the multilevel statistical approach gives a clearer interpretation of the multilevel data. The aim of this method is to analyze the original dataset in three different parts: offset, between-batch, and within-batch variation.

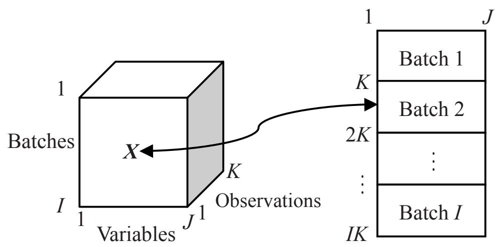

Fig.1 Structure of the dataset of a batch process

MLICA-PCA

In this section, a two-level case will be discussed to explain the method. Any extension to more levels is a straightforward process and needs only to follow the same approach. MLICA-PCA is an extension of ICA-PCA that can be used if the variation in the data occurs simultaneously on different levels. The ICA-PCA method mentioned in Section 2 is a combination of ICA and PCA, which can extract full information, both non-Gaussian and Gaussian, from the process. However, when the ICA-PCA method is applied to the dataset $\boldsymbol{X}$ represented in Fig.1, the between-batch variation and the within-batch variation are confounded, which seriously compromises the interpretation of variation in the data.

Multilevel based models contain sub-models for the different types of variation in the data. For batch processes, these are the between-batch and the within-batch sub-models. The decomposition of the dataset $\boldsymbol{X}$ consists of a mean-centering of each matrix $\boldsymbol{X}_i$, where $\boldsymbol{X}_i$ is the $K\times J$ partition of matrix $\boldsymbol{X}$ belonging to batch i (Jansen *et al.*, 2005):

$$\boldsymbol{X}_{c,i} = \boldsymbol{X}_i - \boldsymbol{1}_K \boldsymbol{m}_i^{\mathrm{T}}, \quad (6)$$

where $\boldsymbol{m}_i^{\mathrm{T}}$ is a row vector containing the column means of matrix $\boldsymbol{X}_i$, and $\boldsymbol{X}_{c,i}$ is a $K\times J$ matrix containing only the dynamic within-batch variation belonging to batch i. The stationary differences between the batches are described in matrix $\boldsymbol{M}$, which is concatenated by the I vectors $\boldsymbol{m}_i^{\mathrm{T}}$. The between-batch sub-model can be determined by performing ICA-PCA on $\boldsymbol{M}$,

$$\boldsymbol{M}^{\mathrm{T}} = \boldsymbol{A}_{\mathrm{b}}\boldsymbol{S}_{\mathrm{b}} + \boldsymbol{E}_{\mathrm{b}}, \quad (7)$$

$$\boldsymbol{E}_{\mathrm{b}}^{\mathrm{T}} = \boldsymbol{T}_{\mathrm{b}}\boldsymbol{P}_{\mathrm{b}}^{\mathrm{T}} + \boldsymbol{F}_{\mathrm{b}}, \quad (8)$$

where $\boldsymbol{S}_{\mathrm{b}}$ contains independent components of between-batch variation, $\boldsymbol{T}_{\mathrm{b}}$ is the between-batch score matrix, and $\boldsymbol{F}_{\mathrm{b}}$ is an $I\times J$ matrix containing the residuals of the between-batch model.

Similarly, the within-batch sub-model can be calculated by fitting an ICA-PCA model on $\boldsymbol{X}_{\mathrm{c}}$, which is given in the following equations:

$$\boldsymbol{X}_{\mathrm{c}}^{\mathrm{T}} = \boldsymbol{A}_{\mathrm{w}}\boldsymbol{S}_{\mathrm{w}} + \boldsymbol{E}_{\mathrm{w}}, \quad (9)$$

$$\boldsymbol{E}_{\mathrm{w}}^{\mathrm{T}} = \boldsymbol{T}_{\mathrm{w}}\boldsymbol{P}_{\mathrm{w}}^{\mathrm{T}} + \boldsymbol{F}_{\mathrm{w}}, \quad (10)$$

where $\boldsymbol{X}_{\mathrm{c}}$ is an $IK\times J$ matrix which is concatenated by all matrices $\boldsymbol{X}_{c,i}$, $\boldsymbol{S}_{\mathrm{w}}$ contains independent components of all the within-batch variation $S_{\mathrm{w},i}$ (independent components that belong to batch i), $\boldsymbol{T}_{\mathrm{w}}$ contains all the within-batch scores $T_{\mathrm{w},i}$ (that belong to batch i), and $\boldsymbol{F}_{\mathrm{w}}$ is an $IK\times J$ matrix containing the residuals of the within-batch model.

The overall MLICA-PCA model is divided into several sub-models. In the case of two levels used in this paper, the overall model is given by

$$\boldsymbol{X} = mean(\boldsymbol{X}) + [(\boldsymbol{A}_{\mathrm{b}}\boldsymbol{S}_{\mathrm{b}})^{\mathrm{T}} + \boldsymbol{T}_{\mathrm{b}}\boldsymbol{P}_{\mathrm{b}}^{\mathrm{T}}] + [(\boldsymbol{A}_{\mathrm{w}}\boldsymbol{S}_{\mathrm{w}})^{\mathrm{T}} + \boldsymbol{T}_{\mathrm{w}}\boldsymbol{P}_{\mathrm{w}}^{\mathrm{T}}] + (\boldsymbol{F}_{\mathrm{b}} + \boldsymbol{F}_{\mathrm{w}}), \quad (11)$$

where $mean(\boldsymbol{X})$ is the overall mean of matrix $\boldsymbol{X}$, the second part of the right-hand side of Eq.(11) models the between-batch variation of the process, and the third part models the within-batch variation. The residuals of the overall model are given in the fourth part, which contains the between-batch residual $\boldsymbol{F}_b$ and the within-batch residual $\boldsymbol{F}_w$.

Comparing Eq.(11) with Eqs.(1) and (2), the multilevel model can be regarded as a constrained version of the ICA and PCA models. Hence, the ICA and PCA models will explain as much or more variation in the data than a multilevel model in which an equal number of components are fitted. The more constrained the multilevel model is, the less variation in the data it will explain for an equal number of fitted components (Jansen *et al.*, 2005). However, different types of variation in the data are described by different sub-models, giving the multilevel model a much better interpretability of multilevel data. Therefore, the trade-off is between reduced accuracy and increased interpretability. When the data have no multilevel feature, the multilevel model is equal to the ICA and PCA models. Since the batch process data are decomposed into two different parts, some variables play an important role in the between-batch part, some are important to the within-batch part and others may be important to both parts. Compared to conventional methods such as MPCA, fault diagnosis and its interpretation become easier using MLICA-PCA. In addition, the non-Gaussian and Gaussian information are monitored separately, providing another way to simplify fault diagnosis and interpretation.

BATCH PROCESS MONITORING STRATEGY BASED ON MLICA-PCA

Given the centered data matrix $\boldsymbol{X}$, two sub-models are built. One models the between-batch variation of the process, and the other models the within-batch variation, as shown in Eqs.(7)~(10). Since the monitoring strategy based on these two sub-models are similar, we demonstrate them in a unified form in this section. Suppose r independent components are extracted, $\boldsymbol{s}=[s_1, s_2, \cdots, s_r]^{\mathrm{T}}\in\mathbb{R}^{r\times 1}$. To monitor the non-Gaussian part of the process, I^2, a new statistic, is defined as (Lee *et al.*, 2004)

$$I^2 = \boldsymbol{s}^{\mathrm{T}}\boldsymbol{s}. \tag{12}$$

After the non-Gaussian information has been extracted, the residual matrix $\boldsymbol{E}$ is obtained. PCA is used to analyze it, expanding $\boldsymbol{E}$ as follows:

$$\boldsymbol{E} = \sum_{i=1}^{k} t_i \boldsymbol{p}_i^{\mathrm{T}} + \boldsymbol{F}, \tag{13}$$

where $\boldsymbol{F}$ is the residual resulting from the PCA model, t_i is the ith principal component. The limits of T^2 and SPE statistics to monitor the remaining Gaussian part of the process are defined below (Nomikos and MacGregor, 1994):

$$T^2 = \sum_{i=1}^{k} \frac{t_i^2}{\lambda_i} \le \frac{k(n-1)}{n-k} F_{k,(n-k),\alpha}, \tag{14}$$

$$SPE = \boldsymbol{f}\boldsymbol{f}^{\mathrm{T}} = \boldsymbol{e}(\boldsymbol{I}-\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}})\boldsymbol{e}^{\mathrm{T}} \le SPE_\alpha, \tag{15}$$

$$SPE_\alpha = \theta_1 \cdot \left[1 + \frac{c_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right]^{1/h_0}, \tag{16}$$

where k is the number of principle components; $\theta_i = \sum_{j=k+1}^{m} \lambda_j^i$ for i=1, 2, 3; $h_0 = 1 - 2\theta_1\theta_3/(3\theta_2^2)$; α is the significance level; c_α is the normal deviation corresponding to the upper $1-\alpha$ percentile (Nomikos and MacGregor, 1994).

In PCA monitoring, the confidence limit is based on a specified distribution as shown in Eqs.(14)~(16) based on 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 statistics cannot be determined directly from a particular approximate distribution. An alternative approach to defining the nominal operating region of the I^2 statistic is to use the kernel density estimation (KDE) (Chen *et al.*, 2004).

The implementation of the monitoring method as mentioned above consists of two procedures: off-line modeling and on-line monitoring. In the off-line modeling procedure, two ICA-PCA monitoring sub-models are developed in the normal operating condition (NOC). Fault detection and isolation are executed using these monitoring sub-models in the on-line monitoring procedure. The algorithm flows of the two monitoring procedures are summarized in Figs.2 and 3, respectively.

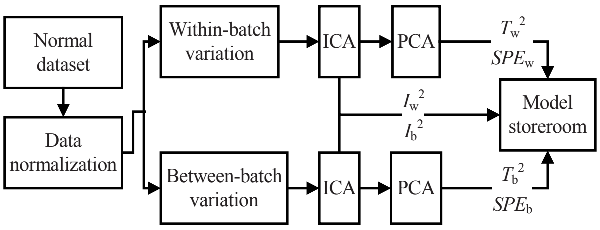

Fig.2 Flow chart of off-line modeling procedure

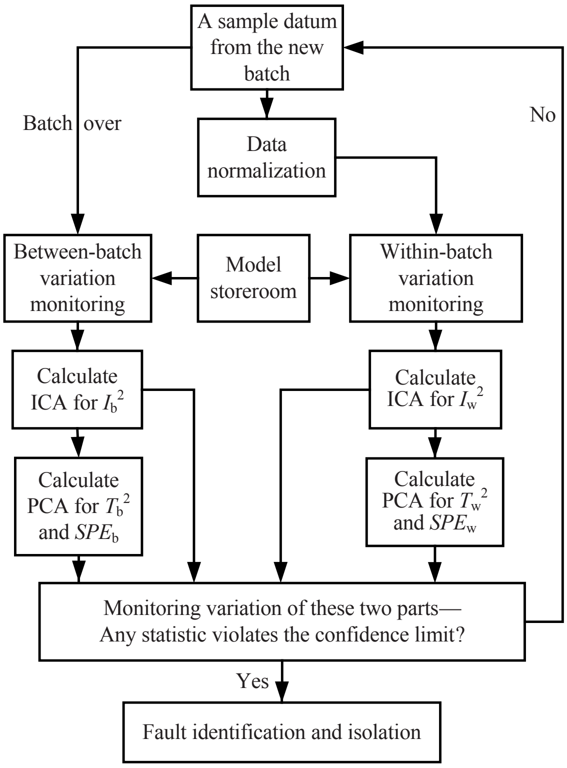

Fig.3 On-line batch process monitoring strategy based on MLICA-PCA

SIMULATION AND RESUSLTS

The Dupont benchmark data for an industrial batch polymerization were taken from (Nomikos and MacGregor, 1995a). The batch reactor consists of two stages. Approximately 2 h are needed to finish one batch run. However, the critical property measurements are usually taken 12 h or more after finishing each batch run. For this type of problem, it is important to develop an on-line batch process monitoring strategy. More information on the Dupont system can be found in (Nomikos and MacGregor, 1995a).

The dataset comprises 55 batches, in which 36 batches are normal and 19 batches are considered abnormal. The 36 successful batches are used for constructing the monitoring model. Each batch run has a duration of 100 time intervals. Ten variables are measured in the process (Table 1). Because raw material is added in the first stage and its flow rate remains zero during the reaction, we excluded variables 5 and 10 from the monitoring variables. To test the performance of the new monitoring method, several batches with a product that was out of the specification were tested.

Table 1 Measured variables in the Dupont batch process

No.	Measured variable
1	Reactor temperature (T_{R1})
2	Reactor temperature (T_{R2})
3	Reactor temperature (T_{R3})
4	Reactor pressure (p_{R1})
5	Material flow rate (f_{M1})
6	Heating-cooling medium temperature (T_{M1})
7	Heating-cooling medium temperature (T_{M2})
8	Reactor pressure (p_{R2})
9	Reactor pressure (p_{R3})
10	Material flow rate (f_{M2})

First of all, the reference dataset $\boldsymbol{X}$ was used to develop the MLICA-PCA model. $\boldsymbol{X}$ was arranged as a three-way matrix $\boldsymbol{X}_{I\times J\times K}$. Following the off-line monitoring procedure, an MLICA-PCA model was built. We chose four independent components for both the within-batch and between-batch ICA models, according to the negentropy method (Hyvarinen and Oja, 2000). After the non-Gaussian information was extracted from the process, PCA was carried out to analyze the residual Gaussian information. Four principal components were chosen to build the PCA model, which explained about 85% of the total variability in both levels of the datasets. The loadings of the between-batch and within-batch models can be compared to identify the variables that are important to describe the variation between and within the batches. In the ICA model, the loading matrices correspond to the separating matrix. The loadings of the components of the within-batch model are given in Fig.4a. Because each independent component is considered as having an equal role in the ICA model, the mean loadings of all the independent components are used for variable identification. For the PCA model, the loadings of the first principal component are used. The loadings of the components of the between-batch model are given in Fig.4b. Comparing both of the loadings shown in Fig.4, many variables are important to describe the variation in both the within-batch and between-batch levels (variables 1, 2, 4 and 9). However, there are variables that have a high loading only for the between-batch model (e.g., variable 3). This means that these variables vary between the batches, but are relatively constant within each batch. It can be concluded that a relatively large variation in these variables for a new batch indicates abnormal behavior. On the other hand, there are also variables that have a high loading only for the within-batch model (variables 6, 7, and 8). In addition, the non-Gaussian and Gaussian information are monitored separately. Some variables play important roles in the non-Gaussian part of the process, while some are important in the Gaussian part. However, this conclusion could not be drawn from an interpretation of a PCA model of the batch process dataset, since in this model the between-batch variation and within-batch variation are confounded.

For on-line process monitoring, several batches were tested, including normal and abnormal batches. Fig.5 shows the within-batch monitoring results from the normal batch using MLICA-PCA. For comparison, the monitoring results using MPCA are also shown. The dashed lines represent the 99% confidence limits of each statistic. Using the MLICA-PCA model, this batch stayed below the confidence limit for each case, indicating that this batch behaved normally throughout the batch run.

Two examples of on-line monitoring of two abnormal batches are shown in Figs.6 and 7. Fig.6 shows the within-batch variation of the 47th batch in the dataset, and Fig.7 corresponds to the 49th batch. The between-batch monitoring results are shown in Fig.8, including 36 normal and two abnormal batches. The dotted lines represent 99% confidence limits of each statistic. From the within-batch monitoring results shown in Fig.6, we can infer that the fault may have occurred in the first stage. After about 25 intervals, Fig.6b shows the process reverted to normal,

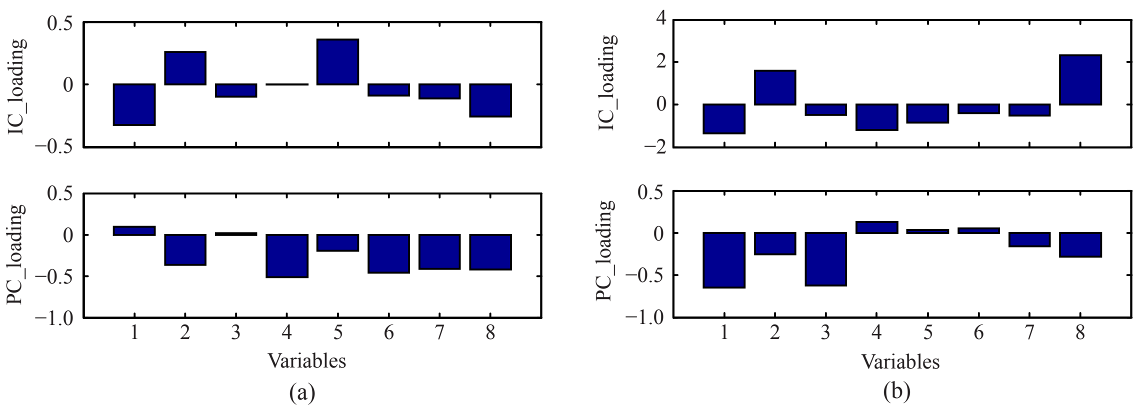

Fig.4 Variable loadings of the components. (a) Within-batch model; (b) Between-batch model

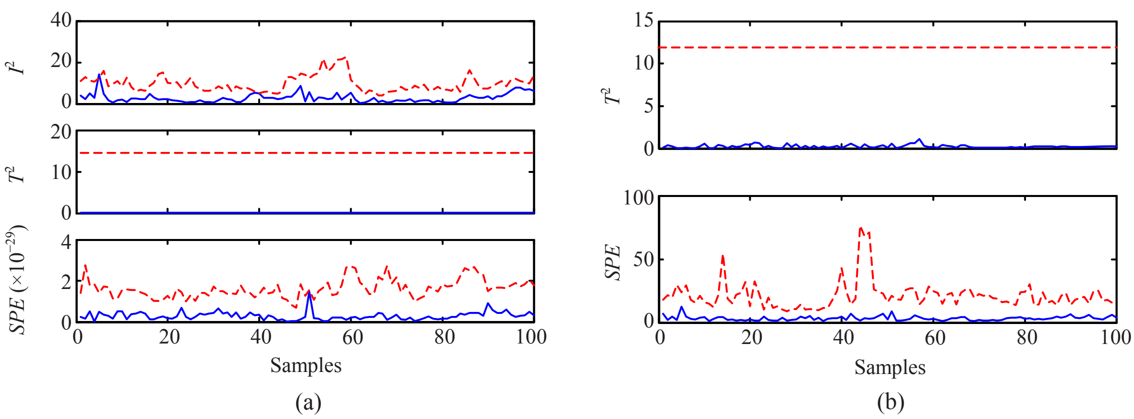

Fig.5 Within-batch monitoring results of the normal batch. The solid lines represent the statistic values of the monitoring batch, and the dashed lines represent their corresponding 99% confidence limits. (a) MLICA-PCA; (b) MPCA

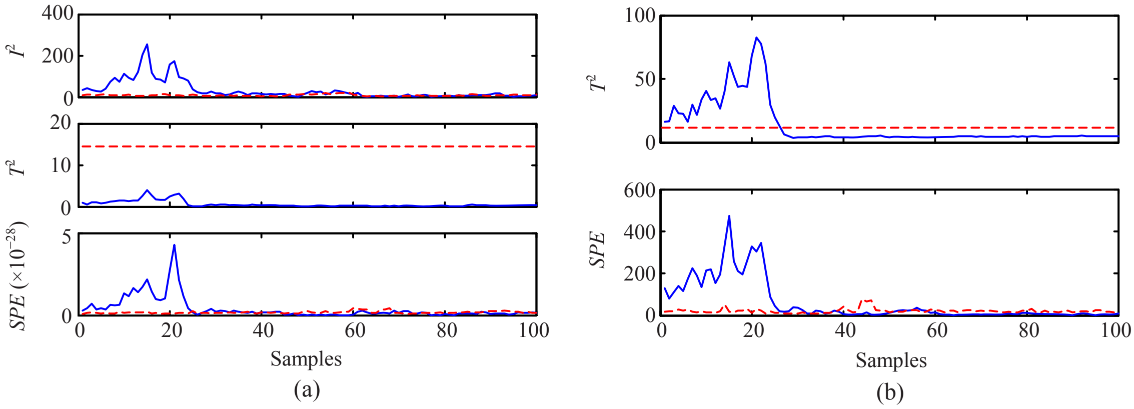

Fig.6 Within-batch monitoring results of batch 47. The solid lines represent the statistic values of the monitoring batch, and the dashed lines represent their corresponding 99% confidence limits. (a) MLICA-PCA; (b) MPCA

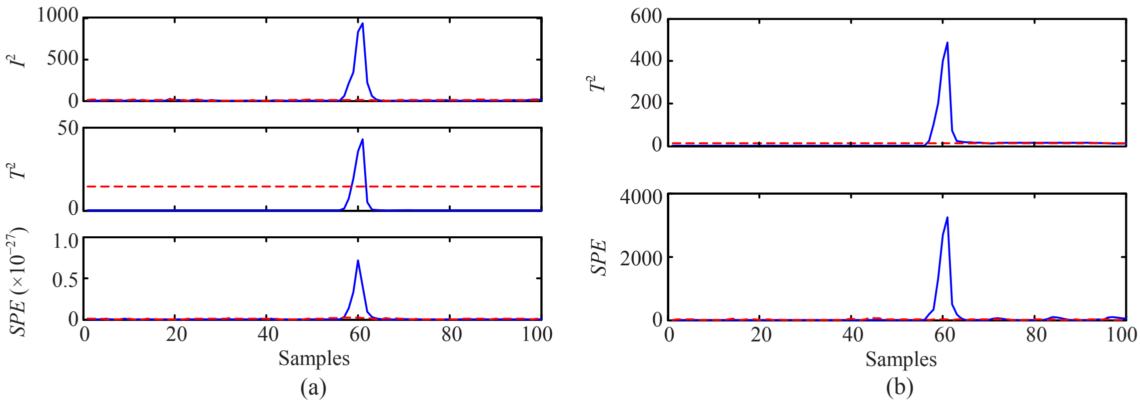

Fig.7 Within-batch monitoring results of batch 49. The solid lines represent the statistic values of the monitoring batch, and the dashed lines represent their corresponding 99% confidence limits. (a) MLICA-PCA; (b) MPCA

since every statistic fell back inside the limit. In this case, if judging the status of the process based on MPCA, a process operator would probably have concluded that a fault had occurred in the process which then corrected itself. However, the fault may have remained in the process, and this batch should have been judged unusual and likely to lead to a low-quality product. During the second stage of the process (Fig.6a), the operation behavior still appeared unusual. Therefore, the operator would have been able to judge the process behavior during the entire batch time, and the batch would have been excluded from the normal ones.

Fig.7 shows batch 49 which was discussed in (Nomikos and MacGregor, 1995a). This batch yielded a product of marginal quality, in that the quality measurement was right at the acceptable limit. The within-batch monitoring results for this batch clearly signalled that something was unusual between time intervals 57 and 65. In spite of its return inside the acceptable control limit, this batch was still characterized as abnormal. Indeed, this deviation did result in a product with a borderline quality. Both methods detected the unusual event. Therefore this batch should also have been excluded from the normal ones.

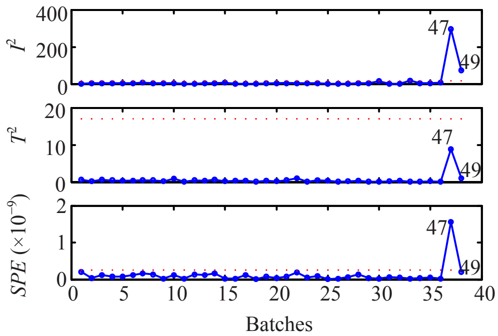

Fig.8 Between-batch monitoring results (36 normal batches and two abnormal batches 47 and 49)

Once a fault or a special event has been detected, it is important to diagnose the event to find an assignable cause. In conventional MPCA, the contribution of each variable to the deviations observed in the statistics can be displayed. They show the group of variables that are primarily responsible for the detected deviations. However, since the within-batch variation and between-batch variation are mixed, and the variables are correlated, the group of responsible variables becomes large, making fault diagnosis difficult. The results of MPCA are given in Fig.9. Fortunately, the proposed MLICA-PCA separates the variation into two parts: within-batch variation and between-batch variation. As discussed, some variables are more important than others in the within-batch part, some are more important in the between-batch part, and some are important in both parts. In Fig.8, the values of two statistics (I_b^2 and SPE_b) of batch 47 are both out of limits, while the T_b^2 statistic is under its limit, indicating that some variables related to the between-batch part are abnormal (both the non-Gaussian and Gaussian). However, the result of batch 49 was different: the non-Gaussian part of the variation was out of limit, while the Gaussian part remained under the limit. This showed that some variables related to the non-Gaussian part of the between-batch variation may become abnormal. For incorporation with the information of the within-batch part, contribution plots were developed for the three statistics in specified intervals (interval 10 for batch 47 and interval 60 for batch 49). In Fig.10a, we can infer that variables 2, 7, 8 and 9 were responsible for the fault, while in Fig.10b, variables 6~9 may have been responsible. The variable loadings in Fig.4 are mostly in accord.

From the preceding examples it is clear that the MLICA-PCA model can monitor two types of variation, and shows good detection performance. Also, the separation of the data level makes fault diagnosis easier, compared to the conventional MPCA method. The new method also monitors the non-Gaussian and Gaussian information. We conclude that the results of the MLICA-PCA model are more interpretable than those from the MPCA model.

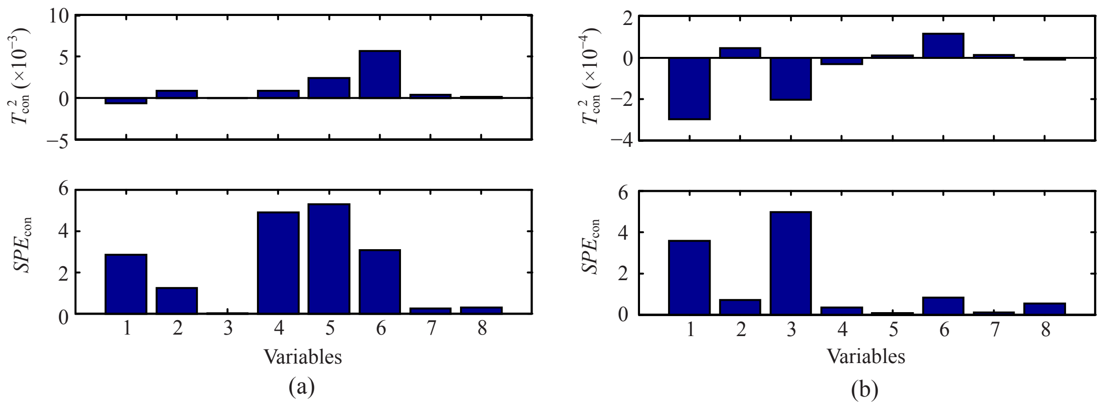

Fig.9 Variable contribution of MPCA. (a) Batch 47; (b) Batch 49

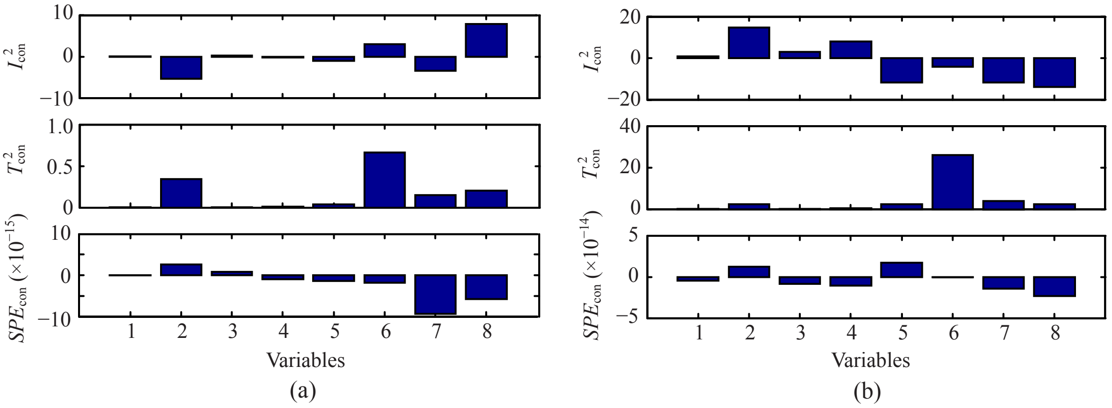

Fig.10 Within-batch variable contribution of MLICA-PCA. (a) Batch 47; (b) Batch 49

CONCLUSION

A novel strategy based on MLICA-PCA has been developed for fault detection and diagnosis. The proposed method provides a separated interpretation for the multilevel batch process data. This reduces the number of variables that might be responsible for faults, making fault diagnosis easier than when using the conventional method. The proposed method extracts the non-Gaussian information from the process, making it possible to monitor the non-Gaussian and Gaussian information separately. The results show the efficiency and importance of the new method for fault diagnosis and location. The proposed method was evaluated in an industrial batch polymerization reactor and showed a superior power of fault detection and diagnosis, compared to the alternative method. Since the proposed method is limited to dealing with the linear nature of a process, further research is needed to extend its application to nonlinear processes.

References

Camacho, J., Picó, J., 2006. Online monitoring of batch processes using multi-phase principal component analysis. *J. Process Control*, **16**(10):1021-1035. [doi:10.1016/j.jprocont.2006.07.005]

Chen, Q., Kruger, U., Leung, A.T.Y., 2004. Regularised kernel density estimation for clustered process data. *Control Eng. Pract.*, **12**(3):267-274. [doi:10.1016/S0967-0661(03)00083-2]

de Noord, O.E., Theobald, E.H., 2005. Multilevel component analysis and multilevel PLS of chemical process data. *J. Chemometr.*, **19**(5-7):301-307. [doi:10.1002/cem.933]

Doan, X.T., Srinivasan, R., 2008. Online monitoring of multi-phase batch processes using phase-based multivariate statistical process control. *Comput. Chem. Eng.*, **32**(1-2):230-243. [doi:10.1016/j.compchemeng.2007.05.010]

Ge, Z.Q., Song, Z.H., 2007. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors. *Ind. Eng. Chem. Res.*, **46**(7):2054-2063. [doi:10.1021/ie061083g]

Hyvarinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. *Neural Network*, **13**(4-5):411-430. [doi:10.1016/S0893-6080(00)00026-5]

Jansen, J.J., Hoefsloot, H.C.J., van der Greef, J., Timmerman, M.E., Smilde, A.K., 2005. Multilevel component analysis of time-resolved metabolic fingerprinting data. *Anal. Chim. Acta*, **530**(2):173-183. [doi:10.1016/j.aca.2004.09.074]

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., Ohno, H., 2003. Monitoring independent components for fault detection. *AIChE J.*, **49**(4):969-976. [doi:10.1002/aic.690490414]

Kiers, H.A.L., Ten Berge, J.M.F., 1994. Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *Br. J. Math. Psychol.*, **47**:109-126.

Lee, J.M., Yoo, C.K., Lee, I.B., 2004. Statistical process monitoring with independent component analysis. *J. Process Control*, **14**(5):467-485. [doi:10.1016/j.jprocont.2003.09.004]

Lee, J.M., Qin, S.J., Lee, I.M., 2006. Fault detection and diagnosis based on modified independent component analysis. *AIChE J.*, **52**(10):3501-3514. [doi:10.1002/aic.10978]

Nomikos, P., MacGregor, J.F., 1994. Monitoring batch processes using multiway principal component analysis. *AIChE J.*, **40**(8):1361-1375. [doi:10.1002/aic.690400809]

Nomikos, P., MacGregor, J.F., 1995a. Multivariate SPC charts for monitoring batch process. *Technometrics*, **37**(1):41-59. [doi:10.2307/1269152]

Nomikos, P., MacGregor, J.F., 1995b. Multi-way partial least square in monitoring batch processes. *Chemometr. Intell. Lab. Syst.*, **30**(1):97-108. [doi:10.1016/0169-7439(95)00043-7]

Simoglou, A., Georgieva, P., Martin, E.B., Morris, A.J., de Azevedo, S.F., 2005. On-line monitoring of a sugar crystallization process. *Comput. Chem. Eng.*, **29**(6):1411-1422. [doi:10.1016/j.compchemeng.2005.02.013]

Timmerman, M.E., Kiers, H.A.L., 2003. Four simultaneous component models for the analysis of multivariate time series for more than one subject to model intraindividual and interindividual differences. *Psychometrika*, **68**(1):105-121. [doi:10.1007/BF02296656]

Yoo, C.K., Lee, J.M., Vanrolleghem, P.A., Lee, I.B., 2004. On-line monitoring of batch processes using multiway independent component analysis. *Chemometr. Intell. Lab. Syst.*, **71**(2):151-163. [doi:10.1016/j.chemolab.2004.02.002]