Multiway Independent Component Analysis Mixture Model and Mutual Information Based Fault Detection and Diagnosis Approach of Multiphase Batch Processes

Jie Yu, Jingyan Chen, and Mudassir M. Rashid
Dept. of Chemical Engineering, McMaster University, Hamilton, ON L8S 4L7, Canada

DOI 10.1002/aic.14051
Published online February 27, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Batch process monitoring is a challenging task, because conventional methods are not well suited to handle the inherent multiphase operation. In this study, a novel multiway independent component analysis (MICA) mixture model and mutual information based fault detection and diagnosis approach is proposed. The multiple operating phases in batch processes are characterized by non-Gaussian independent component mixture models. Then, the posterior probability of the monitored sample is maximized to identify the operating phase that the sample belongs to, and, thus, the localized MICA model is developed for process fault detection. Moreover, the detected faulty samples are projected onto the residual subspace, and the mutual information based non-Gaussian contribution index is established to evaluate the statistical dependency between the projection and the measurement along each process variable. Such contribution index is used to diagnose the major faulty variables responsible for process abnormalities. The effectiveness of the proposed approach is demonstrated using the fed-batch penicillin fermentation process, and the results are compared to those of the multiway principal component analysis mixture model and regular MICA method. The case study demonstrates that the proposed approach is able to detect the abnormal events over different phases as well as diagnose the faulty variables with high accuracy. © 2013 American Institute of Chemical Engineers *AIChE J*, 59: 2761–2779, 2013
Keywords: multiphase batch process, multiway independent component analysis mixture model, mutual information, non-Gaussian contribution index, fault detection, fault diagnosis

Introduction

Batch processes play integral roles in the production of high-quality, low-volume products such as polymers, pharmaceuticals, special chemicals, materials, and microelectronics. Abnormal operating conditions during critical periods of batch processes can have significant impact on the final product quality and production yield. Furthermore, product quality related measurements as key indicators of process performance are often determined off-line after the completion of a batch. If the final product quality does not meet specifications, then it is difficult to diagnose the causes of degraded product quality during the batch operation, and the nonconforming products can be completely wasted.[1,2] Therefore, the effective monitoring of batch processes is very important to detect and diagnose various kinds of faults that can either be resolved in subsequent batches or more preferably corrected prior to the completion of the current batch. The early detection and reliable diagnosis of abnormal events that might lead to deteriorated product quality can ensure safe and profitable operation and enable corrective actions to be taken before the harmful disturbances may ruin multiple batches.[3–7]

Correspondence concerning this article should be addressed to J. Yu at jieyu@mcmaster.ca.

Key characteristics of batch processes such as finite duration, inherent process nonlinearity, underlying time-varying dynamics, batch-to-batch variations, and multiplicity of operating phases pose different challenges on process monitoring and fault diagnosis.[8,9] Mechanistic model based monitoring techniques including Kalman filter and parity space methods have been proposed for process monitoring.[10] However, in-depth process knowledge is required to build quantitative process models, which may not be readily obtained for complex batch processes. In addition to the fundamental model based approaches, multivariate statistical techniques have been widely applied to monitor and diagnose batch processes.[11,12] The most popular multivariate statistical process monitoring (MSPM) methods for batch processes include multiway principal component analysis (MPCA) and multiway partial least squares (MPLS).[13–18] These approaches typically assume that the batch process data follow a multivariate Gaussian distribution approximately, so that the derived confidence limits for monitoring statistics are valid. Such assumption indicates that the MPCA and MPLS methods may not be appropriate for monitoring non-Gaussian batch processes. Although nonlinear monitoring techniques such as multiway kernel PCA and multiway kernel PLS have been used to monitor batch processes, the underlying second-order statistics like covariance and cross-correlation may not capture the non-Gaussian features most efficiently.[19] Alternately, MSPM techniques based on trilinear decompositions of three-dimensional (3-D) data matrices such as parallel factor analysis and Tucker models have been proposed.[20] Nevertheless, the global model across different operating phases is unable to characterize the essentially multiphase operation of batch processes. Aimed at multiplicity of different phases in batch operation, a phase identification method based on the changes in the cross-correlation structure of the data set is developed, so that each individual phase can be modeled separately.[21] Although this approach is designed to capture the process dynamics of different operating phases, the local PCA model still cannot characterize the non-Gaussianity within each phase very well where the measured variables have non-Gaussian distributions and do not satisfy the multivariate Gaussian density function.

More recently, multiway independent component analysis (MICA) and multiway kernel ICA are developed for monitoring nonlinear and non-Gaussian batch processes.[22–26] Based on higher-order statistics, the MICA monitoring method is used to maximize the negentropy index so as to extract non-Gaussian features with statistical independency. Although the negentropy index serves as a quantitative measure of non-Gaussianity of process data, it is not equivalent to the multimodality in multiphase batch operation and, thus, cannot monitor batch processes in an optimal way. In addition to the routine MSPM methods, machine learning based process monitoring techniques have attracted growing attention. Multiway kernel localized Fisher discriminant analysis is applied to detect faults in batch processes using labeled training data with both normal and faulty samples.[27,28] Nevertheless, class labels in the training data set may not always be available in industrial practice and, thus, a preliminary clustering procedure is often required to identify the class labels for both normal and faulty training samples. Support vector machines (SVMs) is another supervised learning technique that has strong nonlinear monitoring capability and also can handle small training data set.[29,30] However, SVM suffers similar issue of requiring class labels for all the training data. To alleviate such requirement, a one-class SVM method is used for process fault detection and diagnosis, where only normal training data is needed.[31] Despite the need of only normal training data, the one-class SVM method does not specifically take into account the multimodality and inherent dynamics of processes.

Meanwhile, multiway Gaussian mixture model based process monitoring method is proposed to characterize the multiphase batch operation using Gaussian mixture models and Bayesian inference strategy.[32] This approach relies on the assumption that each individual operating phase in batch process follows multivariate Gaussian distribution approximately.[33,34] However, due to the inherent process nonlinearity, the operating data from a single phase may be under non-Gaussian distribution.[35] Multiway hidden Markov model has been used for batch process monitoring, but fault diagnosis is achieved through the classification of different fault types.[36] Following a different technical pathway, hidden semi-Markov model (HSMM) is combined with MPCA for process monitoring where HSMM is used to model the multiphase batch operation by representing each phase as a state and then developing localized MPCA models.[37] Although the multiplicity of operating phases can be accounted for using the HSMM technique, the local MPCA models may not extract the non-Gaussian features within each operating phase for reliable fault detection and diagnosis. An alternative type of multiphase batch process monitoring method utilizes singular points to detect phase changes and select phase-specific dynamic PCA models.[38] However, dynamic PCA is also limited to second-order statistics, so that the non-Gaussian features in a local phase cannot be effectively captured for monitoring multiphase batch processes. In a phase-division-based monitoring method, the local variable correlation information is extracted by clustering the loading matrices or regression coefficient matrices of the time-slice PCA or PLS models so as to identify different operating phases.[38] Nevertheless, the phase identification is highly dependent on the clustering procedure, while selecting the number of clusters may not be a trivial task.

In this study, a MICA mixture model and mutual information based approach is developed for fault detection and diagnosis of multiphase batch processes. The ICA mixture model technique is well suited to handle the multiphase process with non-Gaussianity within individual phases and can adaptively extract the non-Gaussian hidden features of different phases for fault detection. Furthermore, mutual information is used to define a new non-Gaussian contribution index by estimating the statistical dependency between the monitored variables and their projections onto the residual subspace of local ICA models for process fault diagnosis. Such contribution index is based on higher-order statistics and, thus, can characterize the non-Gaussian relationship between different variables and phase-dependent feature subspaces for more reliable fault diagnosis of multiphase batch processes. It should be noted that the term of " "non-Gaussianity" in this article is used to describe process measurement data following non-Gaussian instead of multivariate Gaussian distribution.

The remainder of the article is organized as follows. First, a brief review of the conventional MPCA mixture model and MICA based batch process monitoring methods is given. Then, the novel MICA mixture model based fault detection and mutual information based fault diagnosis approach is developed. The case study examples demonstrate the effectiveness of the proposed method through its application to the fed-batch penicillin fermentation process. Finally, the conclusions of this work are summarized.

Preliminaries

MPCA mixture model based batch process monitoring

MPCA mixture model based process monitoring is used to monitor batch processes with multiple operating modes. Consider the batch process data $X \in \mathbb{R}^{I \times J \times K}$ with total I batches, J measurement variables, and K sampling instants, which can be unfolded to form a two-dimensional (2-D) matrix of the form $\tilde{X} \in \mathbb{R}^{I \times JK}$. Then, the variables at each sampling instant corresponding to every column of the unfolded data matrix are mean centered and scaled to remove the average batch trajectory. The preprocessed data are then rearranged to form a matrix of the form $X \in \mathbb{R}^{J \times IK}$. Further, a preliminary clustering procedure is conducted to classify the batch process data into L different operating phases. With the identified phases, the corresponding L localized PCA models can be built to represent the shifting covariance structures. Thus, the classified subset of data from the lth operating phase with n_l samples can be monitored through the localized PCA models.[39,40] Assuming that the first g_l principal components are retained in the lth local PCA model, the

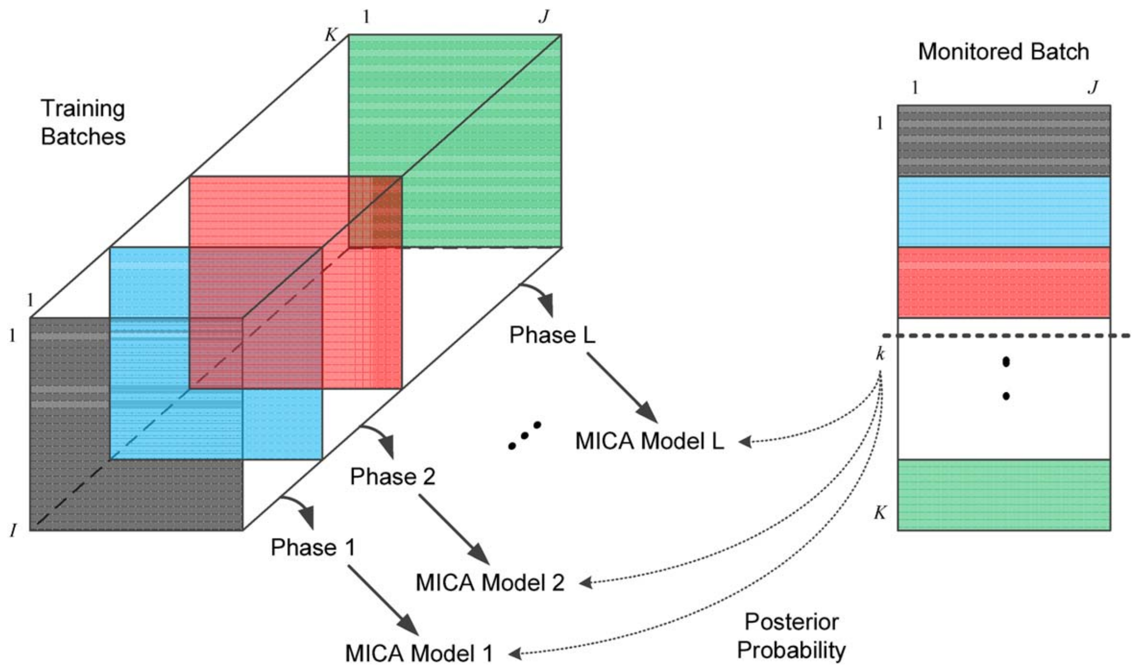

Figure 1. Illustration of the MICA mixture model based fault detection method.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

squared prediction error (SPE) and T^2 indices can be computed as

$$\text{SPE}=\|\left(I-\hat{P}\hat{P}^T\right)x_t(k)\|^2 \tag{1}$$

and

$$T^2=x_t(k)^T\hat{P}\Lambda_g^{-1}\hat{P}^T x_t(k) \tag{2}$$

where $x_t(k)$ is the J-dimensional measurement sample from a test batch at the kth sampling instant, $\hat{P}$ denotes the loading matrix of principal component subspace, and Λ_g represents the matrix of eigenvalues for the retained principal components. Further, the combined SPE/T^2 contribution index within the localized PCA models can be calculated for fault diagnosis.[41]

MICA based batch process monitoring

MICA has been applied to batch process monitoring where the independent components (ICs) extracted from multivariate process measurement data are used to identify the non-Gaussian features for fault detection and diagnosis.[8,22] Consider the batch process data $X\in\mathbb{R}^{I\times J\times K}$, which is unfolded to 2-D matrix $X\in\mathbb{R}^{I\times JK}$. Then, the variables at each sampling instant corresponding to every column of the unfolded data matrix are mean centered and scaled to remove the average batch trajectory. After such data preprocessing, the matrix is fed into MICA algorithm to extract the underlying ICs for process monitoring and fault detection. The measurement matrix $X(k)$ with I batches and J variables at the kth sampling instant can be expressed as a linear combination of M ICs $s_1(k), s_2(k), \ldots, s_M(k)$ as follows

$$X(k)=\sum_{m=1}^{M}a_m(k)s_m(k)+e(k) \tag{3}$$

which can be rewritten in the matrix form

$$X=AS+E \tag{4}$$

where $A=[a_1(k)a_2(k)\ldots a_M(k)]\in\mathbb{R}^{I\times M}$ is the mixing matrix, $S=\left[s_1(k)^T s_2(k)^T\ldots s_m(k)^T\right]^T\in\mathbb{R}^{M\times J}$ is the IC matrix, and $E\in\mathbb{R}^{I\times J}$ is the residual matrix. The objective of ICA algorithm is to calculate a demixing matrix W, so that the components of the following reconstructed data matrix

$$\hat{S}=WX \tag{5}$$

become as independent of each other as possible.

The initial step in the ICA procedure is data whitening, which can eliminate the cross-correlations among different random variables. Consider the J-dimensional measurement vector $x(k)$ from an arbitrary training batch at the kth sampling instant with the corresponding covariance matrix k. The eigenvalue decomposition of k is given by

$$\kappa=U\Lambda U^T \tag{6}$$

and the whitening transformation is expressed as

$$z(k)=Qx(k)=QAs(k)=Bs(k) \tag{7}$$

where $Q=\Lambda^{-1/2}U^T$ and B is an orthogonal matrix. After the whitening transformation, the ICs can be estimated from Eq. 7 as

$$\hat{s}(k)=B^Tz(k)=B^TQx(k) \tag{8}$$

Thus, the demixing matrix W can be obtained as

$$W=B^TQ \tag{9}$$

To calculate B, each column vector $b_i\in B$ is initialized and updated, so that the ith IC $\hat{s}_i=b_i^Tz$ has the maximized non-Gaussianity. Negentropy is a measure of non-Gaussianity that is based on the information theory quantity of

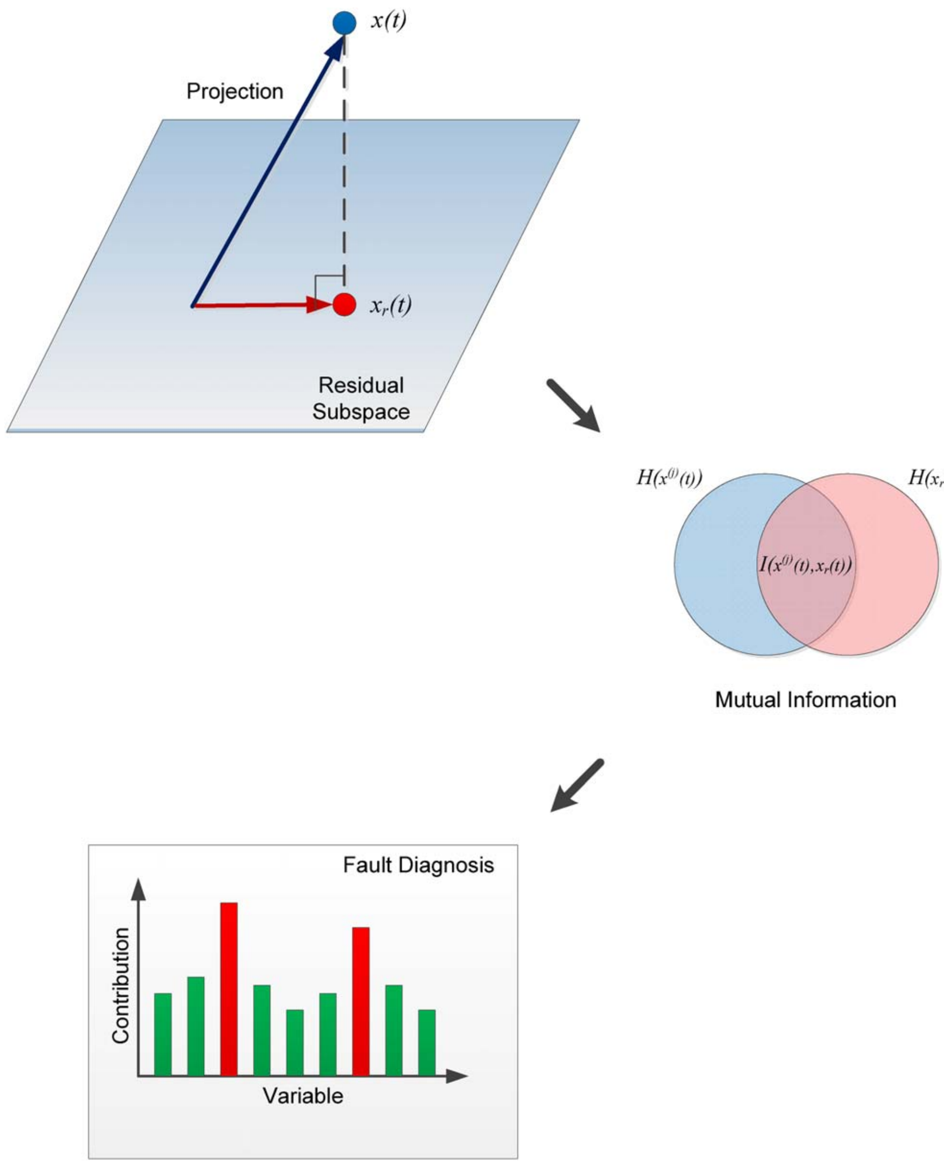

Figure 2. Illustration of the mutual information based fault diagnosis method.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

differential entropy, which is used in the fast and robust fixed-point algorithm for ICA that entails maximizing the negentropy. An approximate form of the negentropy index is used in a simple and highly efficient fixed-point algorithm for ICA.[42,43] Further, the ICs are ordered by sorting the row vectors of demixing matrix W based on their Euclidean norms. The optimal number of ICs is selected by observing the percentages of Euclidean norms to determine the break point between the first few ICs to be retained and the remaining ones. In this way, the percentage norms of the top components that are retained are significantly larger than those of the remaining components.[19]

The ICA-based monitoring statistics for fault detection are then defined as

$$I^2(k)=\hat{s}_t(k)^T\hat{s}_t(k) \tag{10}$$

and

$$\text{SPE}\,(k)=e_t(k)^Te_t(k)=(x_t(k)-\hat{x}_t(k))^T(x_t(k)-\hat{x}_t(k)) \tag{11}$$

where the prediction $\hat{x}_t(k)$ for the test sample $x_t(k)$ at the kth sampling instant can be computed as follows

$$\hat{x}_t(k)=Q^{-1}B\hat{s}_t(k)=Q^{-1}BWx_t(k) \tag{12}$$

Hence, the I^2 statistic is used to detect process faults associated with abnormal variations within the IC subspace, whereas the SPE index is able to capture abnormal events that are within residual subspace. The control limits of both I^2 and SPE indices can be computed through kernel density estimation.[44]

MICA Mixture Model and Mutual Information Based Batch Process Fault Detection and Diagnosis

The conventional MICA based monitoring approach is unable to characterize the multiphase operation of batch processes, because the global MICA method cannot model

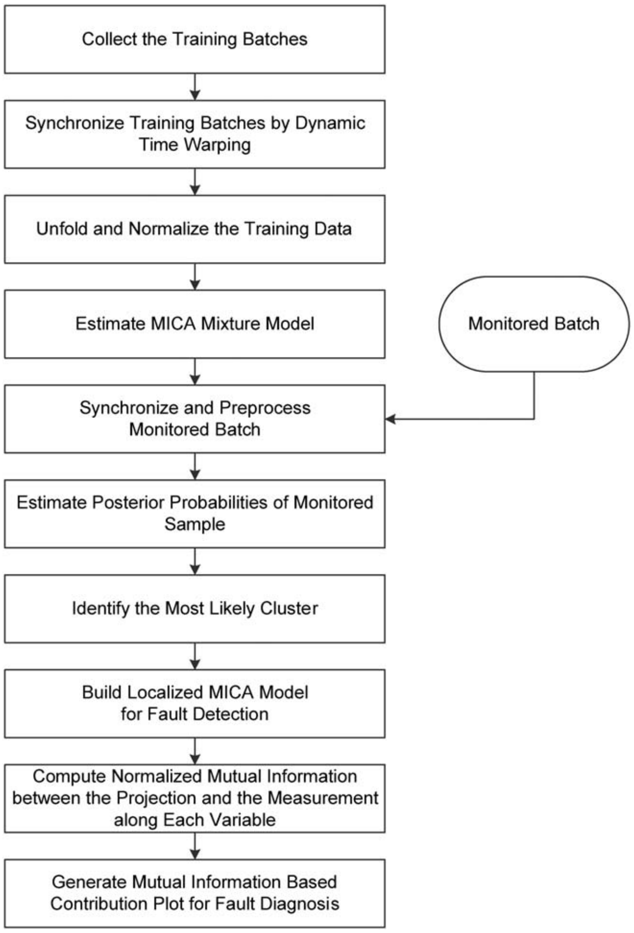

Figure 3. Schematic diagram of the MICA mixture model and mutual information based fault detection and diagnosis approach.

the shifting process dynamics and statistical dependency across various operating phases. Therefore, a novel MICA mixture model based approach is developed where the process observations are modeled as the mixture of a number of

Table 1. Monitored Variables of the Fed-batch Penicillin Fermentation Process

Variable No.	Variable Description
1	Aeration rate (L/h)
2	Agitator power (W)
3	Substrate feed rate (L/h)
4	Substrate feed temperature (K)
5	Substrate concentration (g/L)
6	pH
7	Dissolved oxygen concentration (g/L)
8	Carbon dioxide concentration (g/L)
9	Biomass concentration (g/L)
10	Penicillin concentration (g/L)
11	Fermenter temperature (K)
12	Cooling water flow rate (L/h)
13	Generated heat (kcal)
14	Acid flow rate (L/h)
15	Base flow rate (L/h)
16	Culture volume (L)

mutually exclusive classes that are characterized by a series of non-Gaussian ICA models. Hence, different operating phases can be identified and mapped to multiple localized MICA models. Further, the process faults may be detected within each local MICA model depending on which operating phase the sample instants belong to. After fault detection, the faulty variables of the process upsets can be diagnosed by computing the mutual information based non-Gaussian contribution index between the faulty points and their corresponding projections onto the residual subspace of the local MICA model. In this way, the process variables with the high statistical dependencies between the original measurements and their subspace projections are the most likely abnormal variables.

Batch process data are first expressed as a 3-D matrix $X \in \mathbb{R}^{I \times J \times K}$. In data preprocessing, the unequal durations across different batches can be synchronized through dynamic time warping procedure.[45] Then, the data are unfolded into a 2-D matrix of the form $\tilde{X} \in \mathbb{R}^{I \times JK}$, which is normalized along each column vector. The preprocessed data matrix is further converted into a matrix of the form $X \in \mathbb{R}^{IK \times J}$ through block transpose. The preprocessed data are

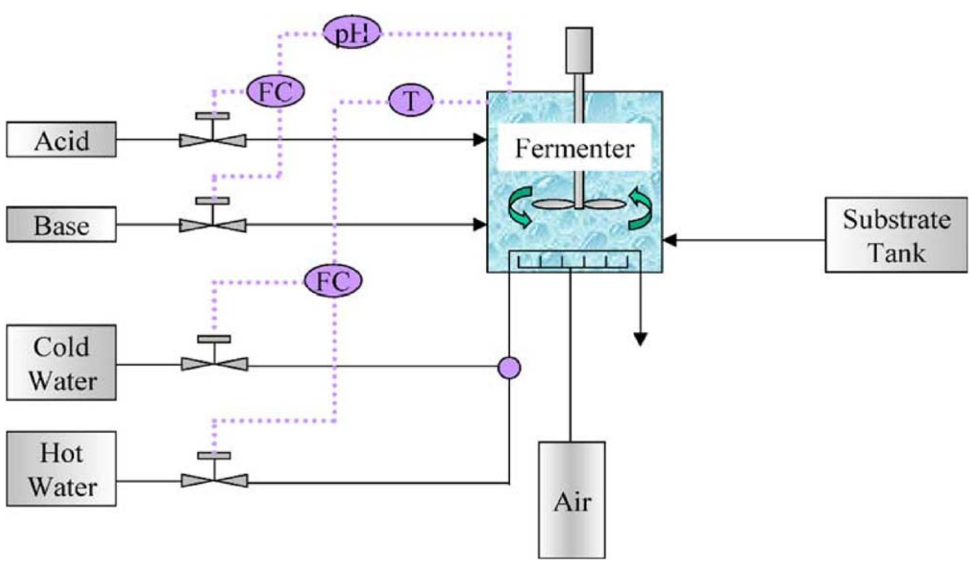

Figure 4. Process flow diagram of the fed-batch penicillin fermentation process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 2. Three Test Cases of the Fed-Batch Penicillin Fermentation Process

Case No.	Test Scenario	Duration (h)
Case 1	Normal operation	0–240
	Drift error in substrate feed rate	240–400
Case 2	Normal operation	0–180
	Step error in aeration rate	180–320
	Normal operation	320–400
Case 3	Normal operation	0–60
	Drift error in substrate feed rate	60–150
	Normal operation	150–320
	Step error in agitator power	320–400

now of the form that is directly applicable to the MICA mixture model method with the rows as the sampling instants and batches whereas the columns as the process variables. The purpose of the above block transpose during data unfolding is to stack different sampling instants with various batches to obtain enough number of unfolded samples for ICA mixture model training.

The ICA mixture model assumes that the observed data in each class are generated by a linear combination of ICs with non-Gaussian probability densities. In addition, the ICs within each ICA class are computed such that the identified latent variables are as statistically independent from each other as possible.[46] The likelihood of the measurement data throughout each batch is given by the following conditional probability density function

$$p(X|\Theta)=\prod_{t=1}^{IK} p(x(t)|\Theta) \tag{13}$$

where $x(t)$ denotes the tth row vector from the unfolded data matrix X and $\Theta=\{\theta_1, \theta_2, \ldots, \theta_L\}$ are the unknown model parameters for different classes $\{C^{(1)}, C^{(2)}, \ldots, C^{(L)}\}$ with L representing the total number of classes. It should be emphasized that the multiple classes in ICA mixture model correspond to the different phases in batch process. The batch measurement sample $x(t)$ is assumed to be generated from a non-Gaussian mixture density function as follows

$$p(x(t)|\Theta)=\sum_{l=1}^{L} p(x(t)|\theta_l)\cdot p(\theta_l) \tag{14}$$

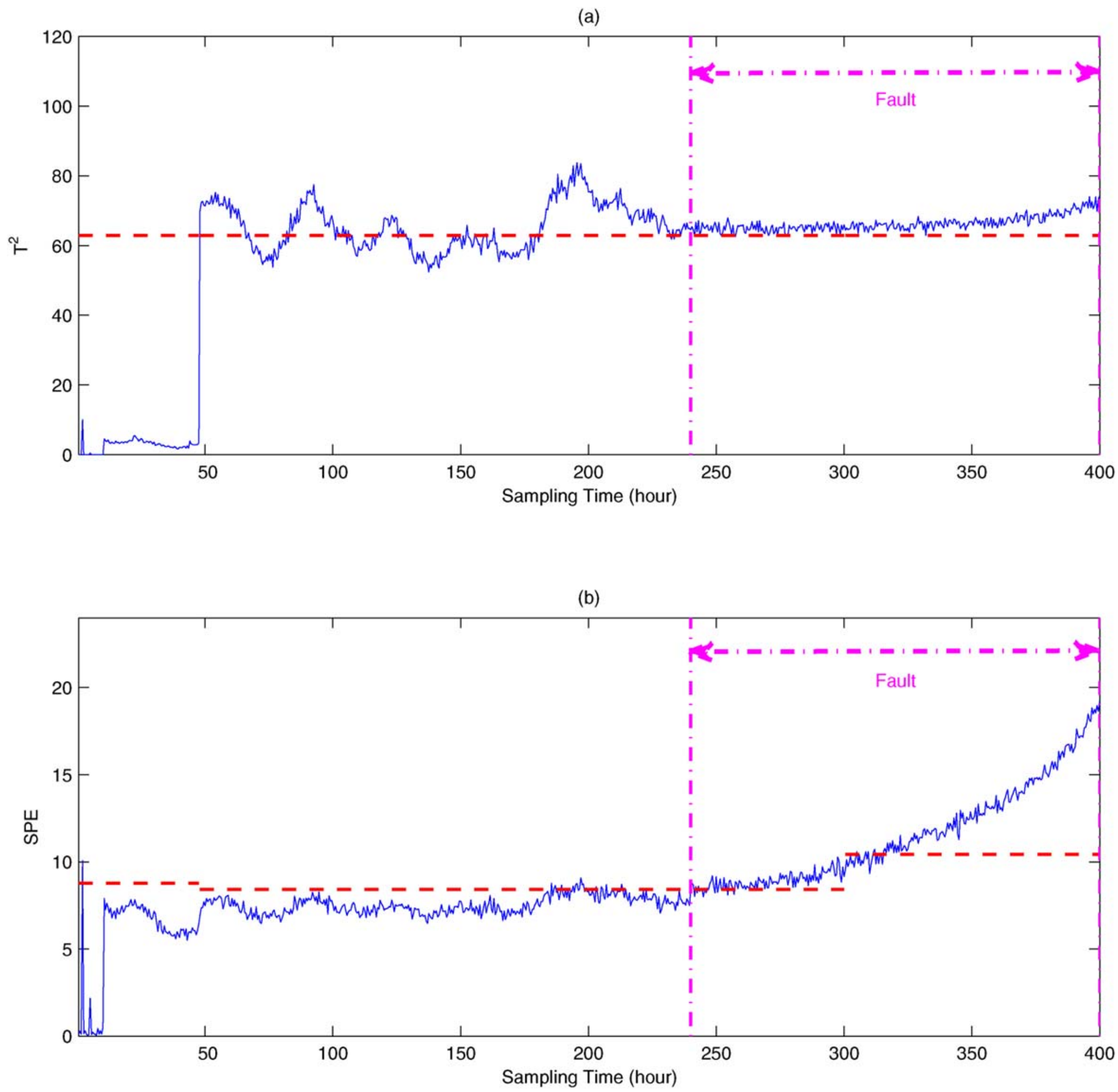

Figure 5. Fault detection results of MPCA mixture model method in the first test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

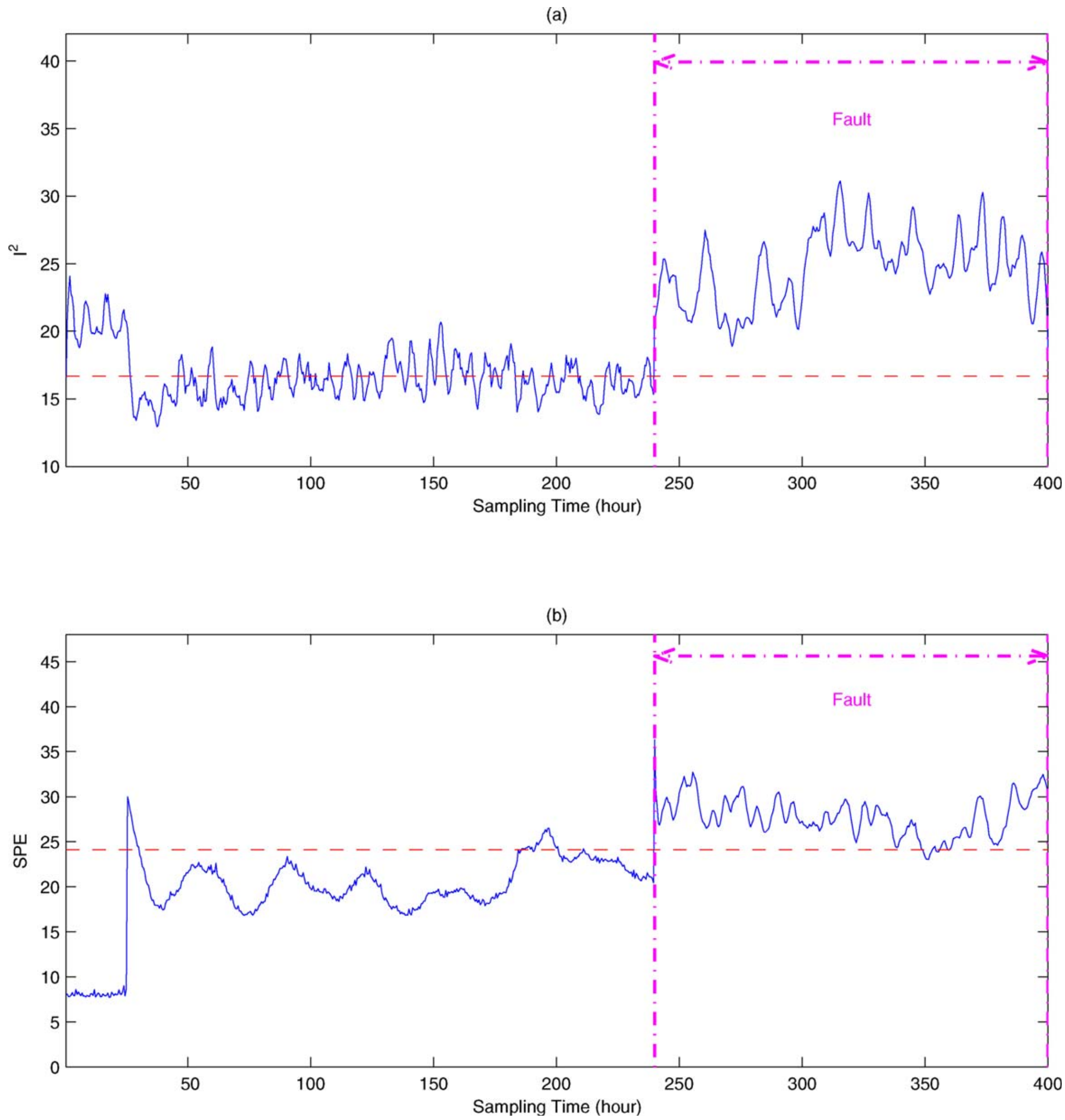

Figure 6. Fault detection results of MICA method in the first test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The probability density function of each component in the mixture model is assumed to be non-Gaussian and the data within each operating phase are expressed as follows

$$x(t)=A_l s_l+e_l \tag{15}$$

where A_l is the mixing matrix, s_l represents IC, and e_l is the bias vector for the lth class.[46]

The gradient ascent based iterative algorithm for learning the ICA mixture model is described as follows:[46]

• Compute the log-likelihood function of the data for each class

$$\log\left[p(x(t)|\theta_l)\right]=\log p(s_l)-\log\left(\det|A_l|\right) \tag{16}$$

where $\theta_l=\{A_l, e_l\}$ and $p(s_l)$ is the multivariate probability density function of the ICs. It can be estimated from Laplacian prior as

$$p(s_l)=\exp\left(-|s_l|\right) \tag{17}$$

The number of classes can be optimized by maximizing the log-likelihood function during the numerical iterations.

• Estimate the posterior probability of each class given the sample vector $x(k)$

$$p\left(C^{(l)}|x(t),\Theta\right)=\frac{p(x(t)|\theta_l)\cdot p(\theta_l)}{\sum_{m=1}^{L} p(x(t)|\theta_m)\cdot p(\theta_m)} \tag{18}$$

where $p(\theta_l)$ is the prior probability of the data from the lth operating phase $C^{(l)}$ and can be estimated as

$$p(\theta_l)=\frac{\sum_{t=1}^{IK} p\left(C^{(l)}|x(t),\Theta\right)}{IK} \tag{19}$$

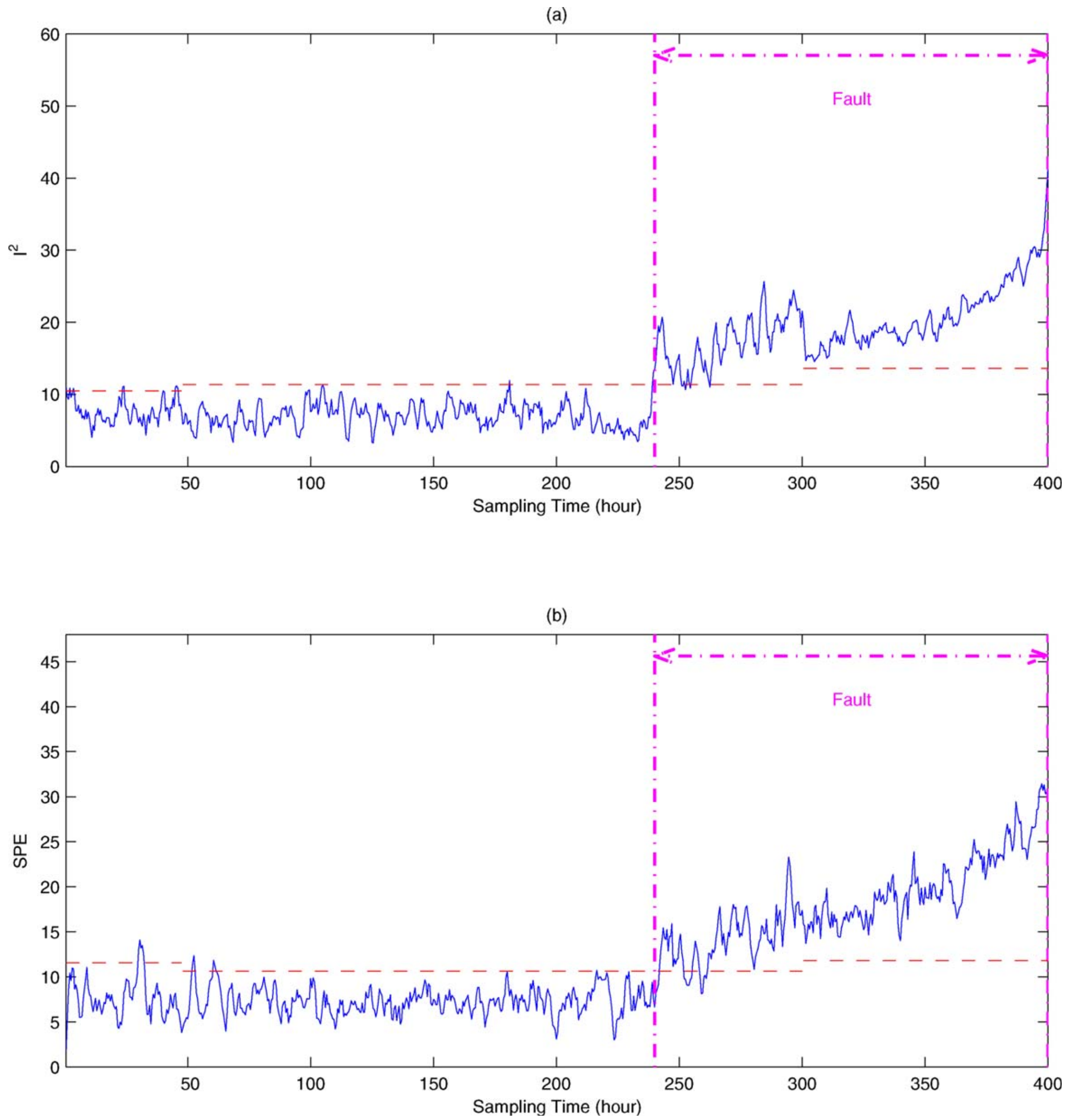

Figure 7. Fault detection results of MICA mixture model based approach in the first test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

• Update the ICA mixing matrix A_l and bias term e_l as follows

$$\Delta A_l = p\left(C^{(l)}|x(t), \Theta\right)\nabla_{A_l}\log p(x(t)|\theta_l) \tag{20}$$

$$e_l = \frac{\sum_{t=1}^{IK} x(t)p(\theta_l|x(t))}{\sum_{t=1}^{IK} p(\theta_l|x(t))} \tag{21}$$

where the gradient of the mixing matrix can be approximated using the ICA model. In the algorithm implementation, the following simplified learning rule of mixing matrix is used

$$\Delta A_l = p\left(C^{(l)}|x(t), \Theta\right)A_l\left[I - \mathrm{sgn}\,(s_l)s_l^T\right] \tag{22}$$

where $\mathrm{sgn}\,(s_l)$ represents the sign function of s_l and I denotes the identity matrix.

The localized MICA model corresponding to each class can, thus, be used for fault detection based on the monitoring statistics that are specific to each batch phase. The posterior probability of the monitored sample belonging to each batch

Table 3. Comparison of Fault Detection Results of the MPCA Mixture Model, MICA, and MICA Mixture Model Based Approaches

Case No.	Monitoring Method	Rate (%)	False Alarm Rate (%)
Case 1	MPCA mixture model	88.16	25.47
	MICA	93.46	14.51
	MICA mixture model	96.57	2.30
Case 2	MPCA mixture model	84.70	2.99
	MICA	92.17	24.57
	MICA mixture model	96.26	2.50
Case 3	MPCA mixture model	86.55	16.16
	MICA	81.14	3.60
	MICA mixture model	95.76	1.86

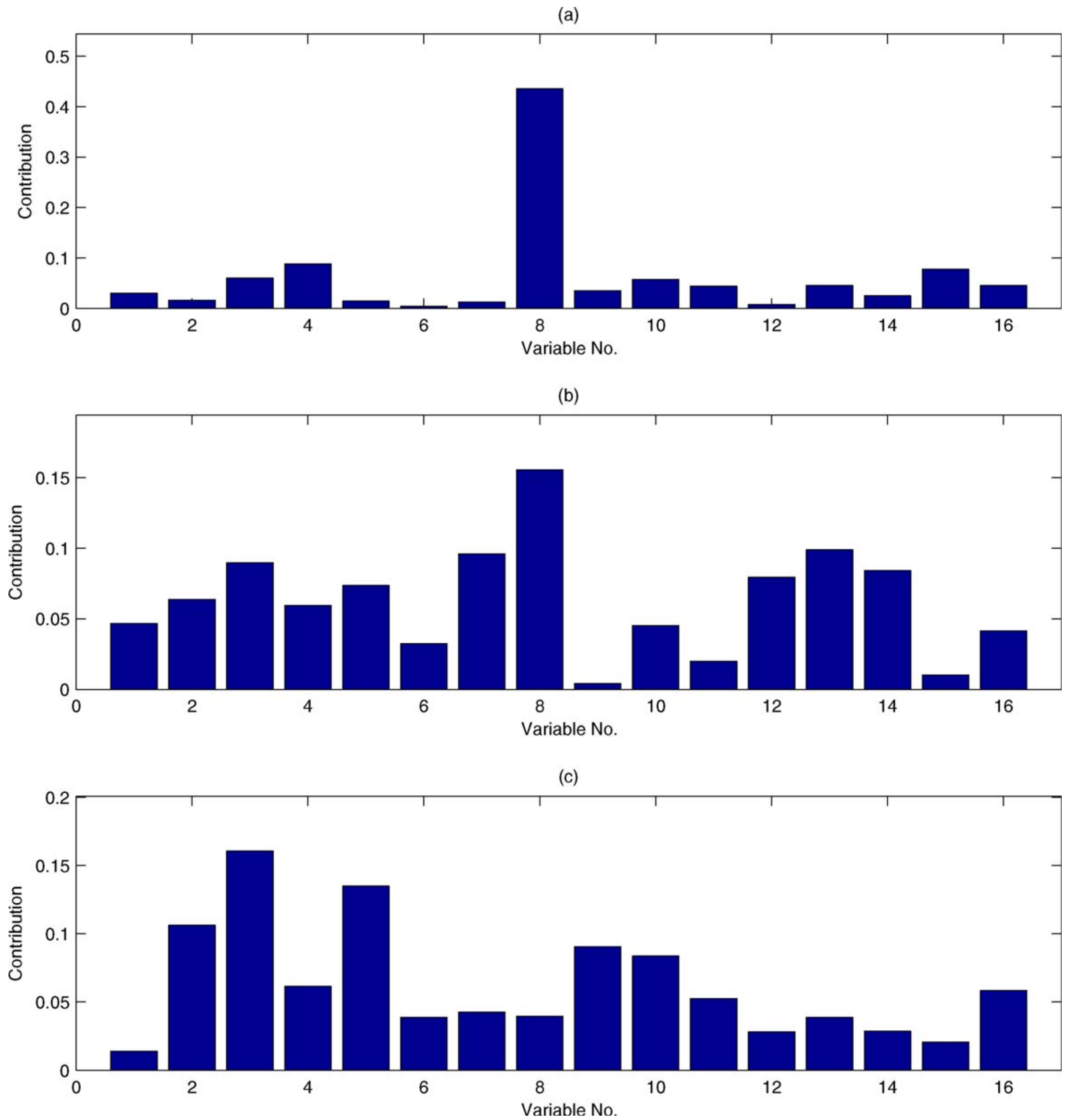

Figure 8. Fault diagnosis results of (a) MPCA mixture model based contribution approach, (b) MICA based contribution approach, and (c) MICA mixture model and mutual information based contribution approach in the first test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

phase can be estimated using Eq. 17 and the operating phase $C^{(M)}$ with the highest posterior probability as follows

$$C^{(M)}=\operatorname{argmax}_{C^{(l)}\in\{C^{(1)},C^{(2)},\ldots,C^{(L)}\}} \quad p\left(C^{(l)}|x(t),\Theta\right) \tag{23}$$

is selected as the particular class for the monitored sample $x(t)$. Then, the localized MICA model for the corresponding operating phase is adopted to compute the monitoring statistics and estimate the control limits for process fault detection. The localized $I^2_{C^{(M)}}$ and $\mathrm{SPE}_{C^{(M)}}$ monitoring statistics for batch operation phase $C^{(M)}$ are computed as

$$I^2_{C^{(M)}}(t)=\hat{s}(t)^T\hat{s}(t) \tag{24}$$

and

$$\mathrm{SPE}_{C^{(M)}}=e(t)^Te(t)=(x(t)-\hat{x}(t))^T(x(t)-\hat{x}(t)) \tag{25}$$

where the projections are given by

$$\hat{s}(t)=B^T_{C^{(M)}}Q_{C^{(M)}}x(t) \tag{26}$$

and

$$\hat{x}(t)=Q^{-1}_{C^{(M)}}B_{C^{(M)}}\hat{s}(t)=Q^{-1}_{C^{(M)}}B_{C^{(M)}}W_{C^{(M)}}x(t) \tag{27}$$

The matrices $Q_{C^{(M)}}$, $B_{C^{(M)}}$, and $W_{C^{(M)}}$ correspond to the localized MICA model for the batch operating phase $C^{(M)}$ that the monitored sample $x(t)$ belongs to. A diagram showing the MICA mixture model based fault detection strategy is given in Figure 1.

After the abnormal events are detected, the faulty process variables need to be further diagnosed, so that the corrective actions can be taken in a timely manner to move the process

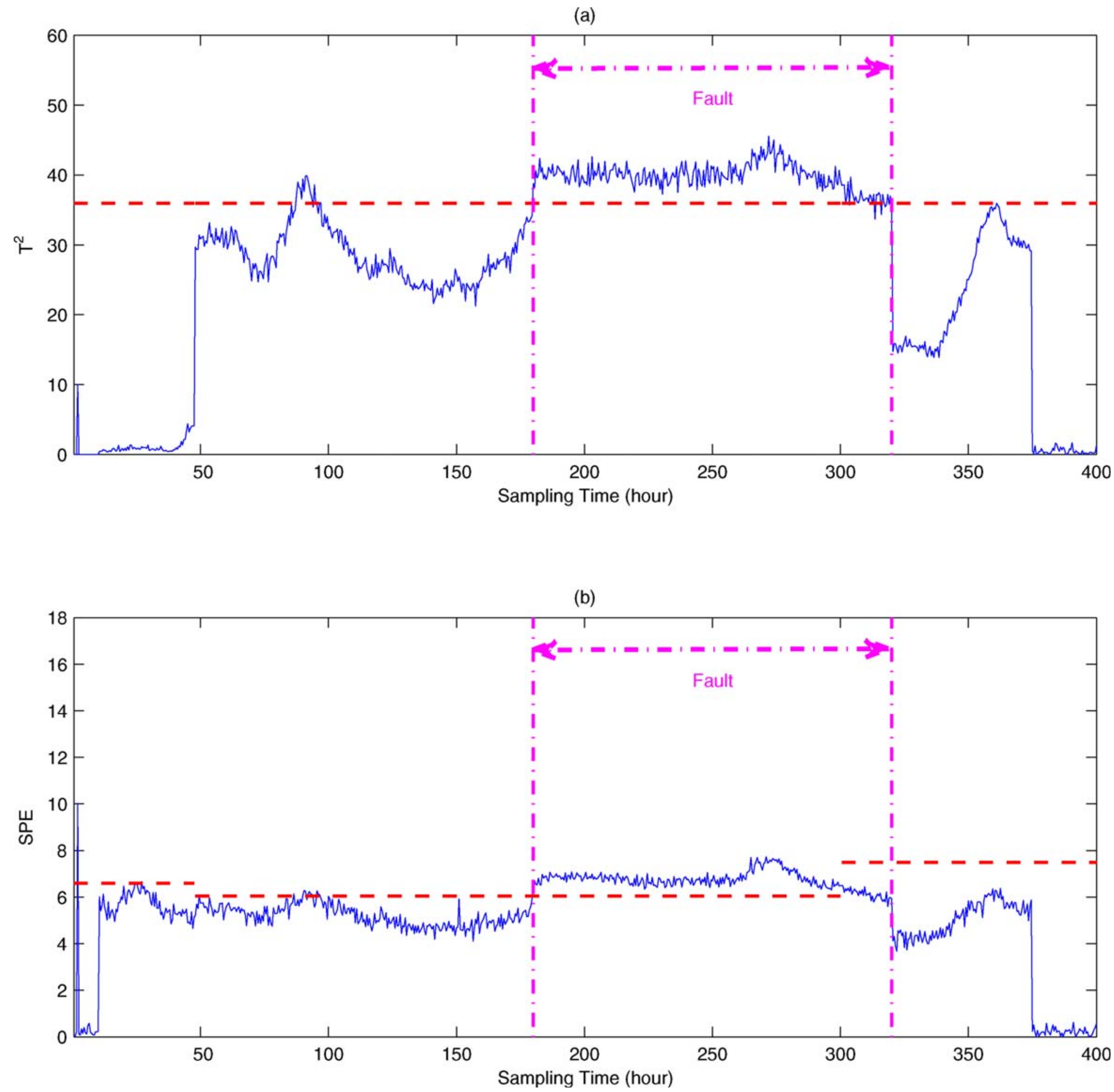

Figure 9. Fault detection results of MPCA mixture model method in the second test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

back to the normal operation. In this study, a mutual information based contribution index is further developed to evaluate the statistical dependency between the original monitored sample and its projection onto the residual subspace of localized ICA model for fault diagnosis. Mutual information is based on the information theory particularly Shannon's entropy. As the measure of system uncertainty for a random variable, Shannon's entropy is defined as

$$H(x) = -\sum_{\alpha=1}^{n} p(x_\alpha)\log p(x_\alpha) \tag{28}$$

where $p(x_\alpha)$ is the probability density function of the possible outcome x_α for a random variable x.[47] The joint entropy of two random variables x and y with possible values x_α and y_β is defined as

$$H(x,y) = -\sum_{\alpha,\beta} p\left(x_\alpha, y_\beta\right)\log p\left(x_\alpha, y_\beta\right) \tag{29}$$

where $p\left(x_\alpha, y_\beta\right)$ is the joint probability density function of the possible outcomes $x=x_\alpha$ and $y=y_\beta$. Then, the mutual information between x and y can be defined as

$$I(x,y) = H(x) + H(y) - H(x,y) = \sum_{\alpha,\beta} p\left(x_\alpha, y_\beta\right)\log\frac{p\left(x_\alpha, y_\beta\right)}{p\left(x_\alpha, y_\beta\right)} \tag{30}$$

The mutual information between the process measurement and its corresponding projection onto the residual subspace of localized MICA model can be normalized and set as a non-Gaussian contribution index for fault diagnosis. The projection of the monitored samples within the residual subspace is computed as

$$x_r(t) = x(t) - \hat{x}(t) = \left(I - Q_{C^{(M)}}^{-1} B_{C^{(M)}} W_{C^{(M)}}\right)x(t) \tag{31}$$

and then the mutual information between the projection $x_r(t)$ and the original measurement along each process variable is computed as follows

$$\zeta^{(j)} = \frac{I\left(x_r(t), x^{(j)}(t)\right)}{H\left(x_r(t), x^{(j)}(t)\right)} \tag{32}$$

where $x^{(j)}(t)$ denotes the measurement value of the monitored sample $x(t)$ on the jth process variable. Further, the

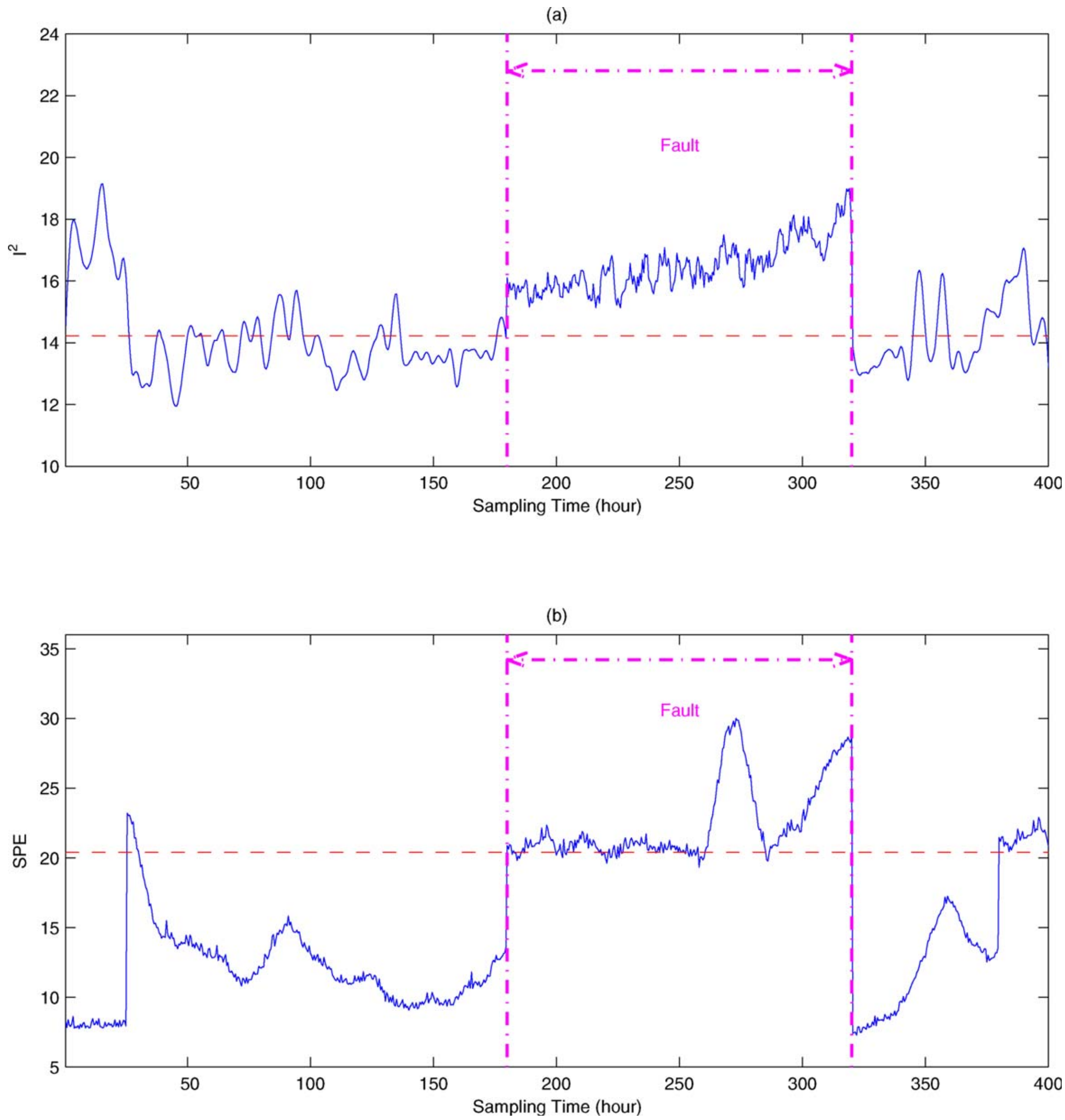

Figure 10. Fault detection results of MICA method in the second test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

above mutual information index can be normalized as follows

$$\zeta_{\mathrm{MI}}(j) = \frac{\zeta^{(j)}(i)}{\sum_{j=1}^{J} \zeta^{(j)}(i)} \tag{33}$$

where J is the total number of process variables. Such normalized mutual information index can be adopted as variable contribution for process fault diagnosis. Typically, the variables with the largest mutual information values are the most likely variables responsible for the process abnormalities. A diagram illustrating the mutual information based fault diagnosis scheme is shown in Figure 2.

The detailed step-by-step procedure for the proposed MICA mixture model and mutual information based fault detection and diagnosis approach of multiphase batch processes is summarized as follows, and the corresponding schematic diagram is shown in Figure 3.

1. Collect training batches of process data for model learning;
2. Synchronize the training batches by dynamic time warping;
3. Unfold and normalize the training data set;
4. Estimate the MICA mixture model using the preprocessed training data and gradient ascent algorithm;
5. Synchronize and preprocess the monitored batch data with unfolding and scaling;
6. Estimate the posterior probabilities of each monitored sample with respect to different operating phases and identify the most likely cluster that the sample belongs to;
7. Select the localized MICA model corresponding to the identified phase and compute the $I^2_{C^{(M)}}$ and $\mathrm{SPE}_{C^{(M)}}$ statistics for fault detection;

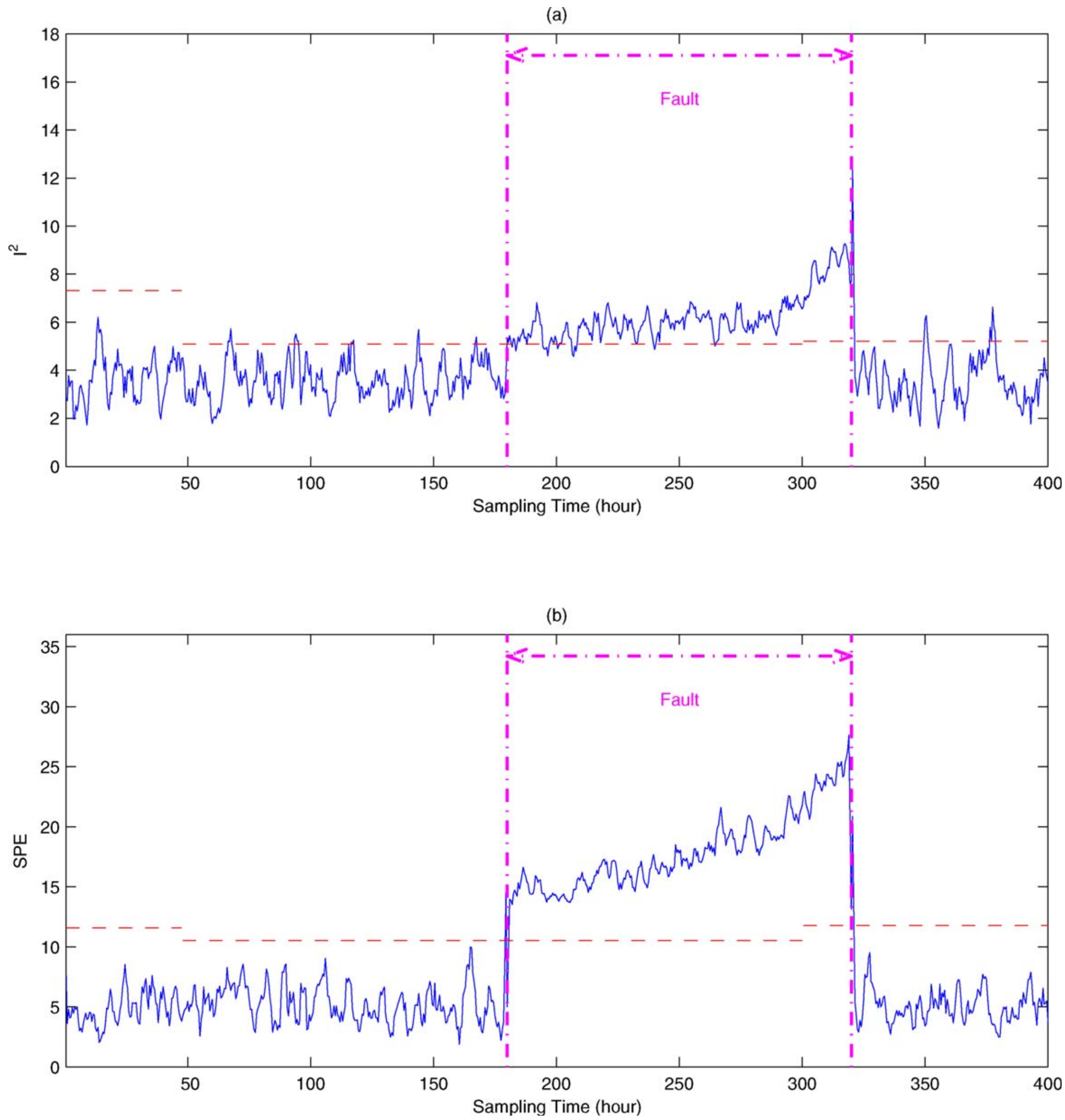

Figure 11. Fault detection results of MICA mixture model based approach in the second test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

8. Project the detected faulty sample onto the residual subspace of the localized MICA model;
9. Compute the normalized mutual information between the projection and the sample measurement along each process variable;
10. Set the normalized mutual information as variable contribution index for fault diagnosis.

Application Example

Fed-batch penicillin fermentation process

The fed-batch penicillin fermentation process is used to evaluate the effectiveness of the proposed MICA mixture model and mutual information based fault detection and diagnosis approach.[48] The process flow diagram of the fed-batch penicillin fermentation process is given in Figure 4. The penicillin fermentation process has three inherent operating phases including lag phase, growth phase, and saturation phase. The penicillin production starts from the growth phase and continues until the end of saturation phase, where a minimum rate of cell growth is needed to maintain a stable penicillin production rate. In addition, this fermentation process begins with batch cultivation and remains for the initial 50 h to promote biomass growth and accumulate adequate cell density for penicillin production. After that, the fermenter was switched to the fed-batch operation with continuous feed of substrate to maintain the cell growth and penicillin production rates. The glucose feed during the fed-batch operation is under open-loop condition, whereas the pH and temperature of the fermenter are controlled by two cascade controllers that manipulate the acid/base equilibrium and the heating/cooling water flow ratio. Moreover, the air is fed into the fermenter to ensure the dissolved oxygen level for cell growth.

In this process, there are totally 16 monitored variables as listed in Table 1. Meanwhile, 20 normal batches with the sampling time of 0.5 h are generated for model training purpose. Three test cases with different types of process faults

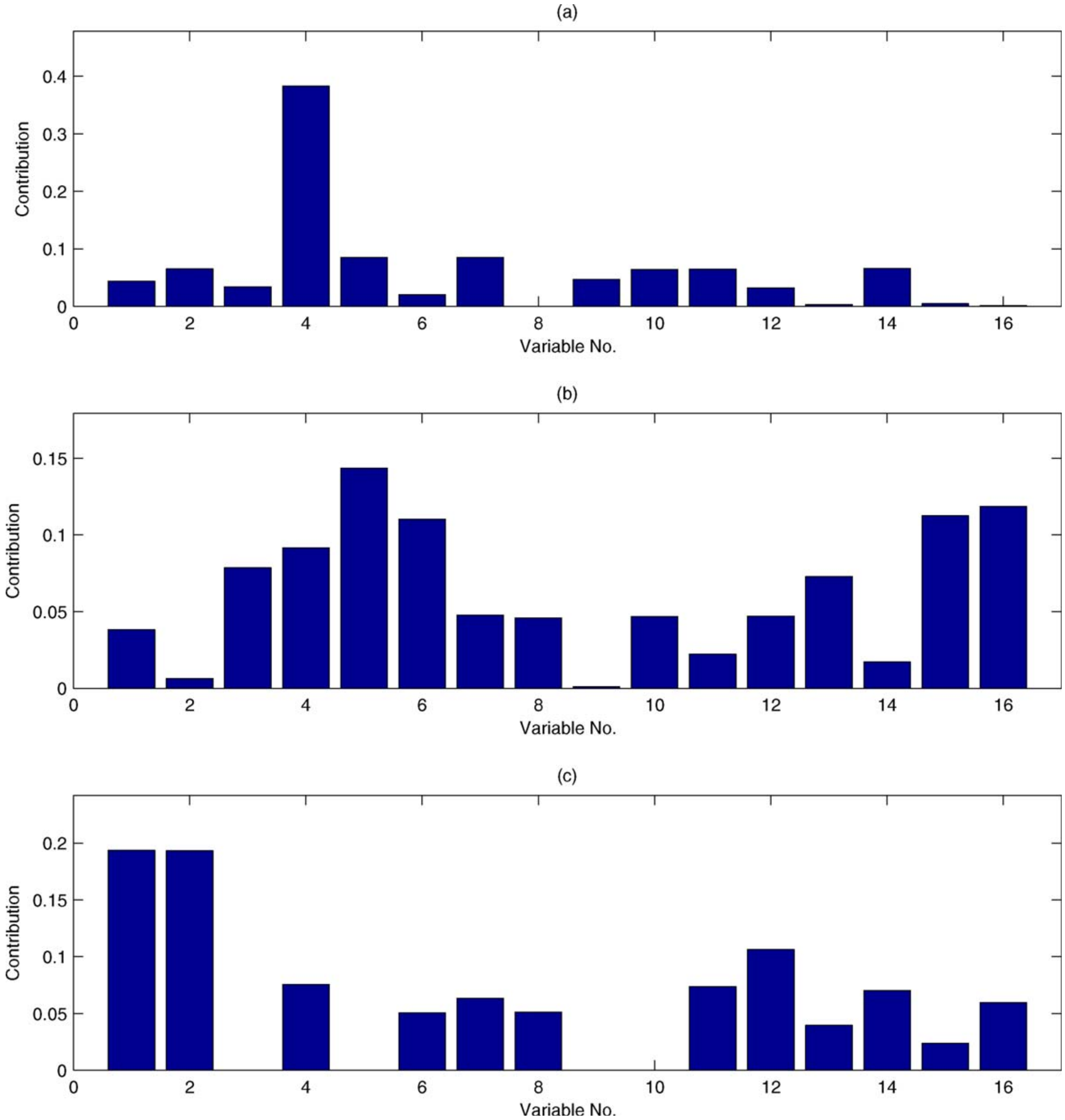

Figure 12. Fault diagnosis results of (a) MPCA mixture model based contribution approach, (b) MICA based contribution approach, and (c) MICA mixture model and mutual information based contribution approach in the second test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

across various operating phases are designed to examine the performance of process monitoring methods, and the faulty scenarios are summarized in Table 2. The first test case starts with normal operation followed by a drift error in the substrate feed rate that is initiated at the 240th hour and lasts until the end of the batch. In the second test case, the normal operation is disturbed by a step error in the aeration rate from the 60th hour and the abnormal event remains for 60 h before the process operation returns to normal condition. A more complex scenario is considered in the third case, where the normal operation is affected by a drift error in the substrate feed rate from the 60th until the 150th hour. Then, the process is back to normal operation and remains until the 320th hour when a step error in the agitator power occurs. In this work, the proposed MICA mixture model based batch process monitoring approach is compared to the MPCA mixture model and conventional MICA based monitoring methods. The confidence level for the control limit estimation in all the methods are set to 9.5%.

Comparison of fault detection and diagnosis results

In the first test case, the process involves a drift error in the substrate feed rate with the duration of 160 h. The fault detection results of the MPCA mixture model method, regular MICA method, and the proposed MICA mixture model based approach are shown in Figures 5–7, respectively. It is seen that the MPCA mixture model based T^2 and SPE statistics result in substantial number of undetected faulty samples and especially false alarms. This approach is unable to extract the non-Gaussian features of the multiphase batch process data, so that its monitoring performance is poor. Meanwhile, it can be observed from Figure 6 that the I^2 index of MICA method does not capture the process fault accurately with significant number of false alarms triggered

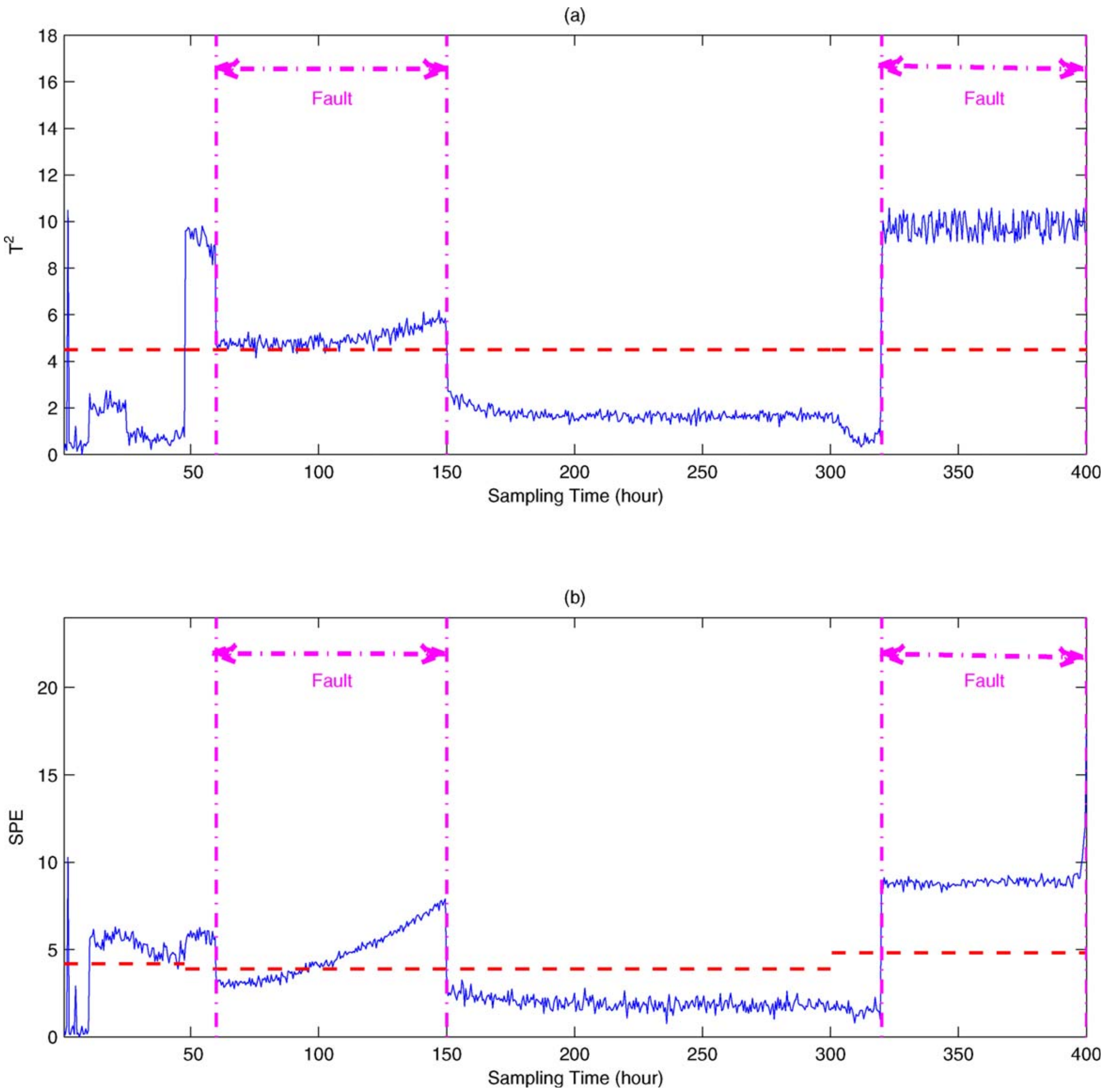

Figure 13. Fault detection results of MPCA mixture model method in the third test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

during normal operation, though the SPE index has improved fault detection capability. In contrast, both the I^2 and SPE indices of the MICA mixture model based monitoring approach can precisely capture the process fault with minimum number of false alarms. The average fault detection and false alarm rates of three different methods are compared in Table 3. One can see that the proposed method leads to the high fault detection rate of 96.57% along with the low false alarm rate of 2.3% only. Although the MPCA mixture model and regular MICA based monitoring methods yield the average fault detection rates of 88.16 and 93.46%, their corresponding average false alarm rates are as high as 25.47 and 14.51%, respectively. Such quantitative comparison indicates that the MPCA mixture model approach cannot extract the key non-Gaussian features for effective process monitoring while the MICA method does not provide reliable monitoring performance due to the multiplicity of operating phases in batch process operation.

After the process fault is detected, fault diagnosis procedure is further conducted to identify the major variables with the most significant upsets. The conventional MPCA mixture model, MICA and the novel MICA mixture model, and mutual information based variable contribution plots are shown in Figures 8a–c, respectively. It should be noted that the MPCA mixture model based contribution plot uses the combined T^2 and SPE contributions. Similarly, the MICA based contribution method combines the I^2 and SPE contributions. From Figures 8a, b, one can see that the variable of CO concentration is identified as the major faulty variable by the MPCA mixture model and MICA based diagnosis methods. Such finding does not coincide with the faulty scenario in the first test case. The MICA mixture model and mutual information based contribution plot, however, correctly identifies substrate feed rate as the major faulty variable. Meanwhile, the abnormality in substrate feed rate can result in the upset of substrate concentration, which has the second largest contribution Figure 8b. The reliable fault diagnosis results of the proposed method are due to two main features: (1) the ICA mixture model can well characterize the multiple operating phases as well as the non-Gaussianity within each phase and (2) the mutual information based contribution index is able to capture the statistical dependency between process variables and non-Gaussian subspaces.

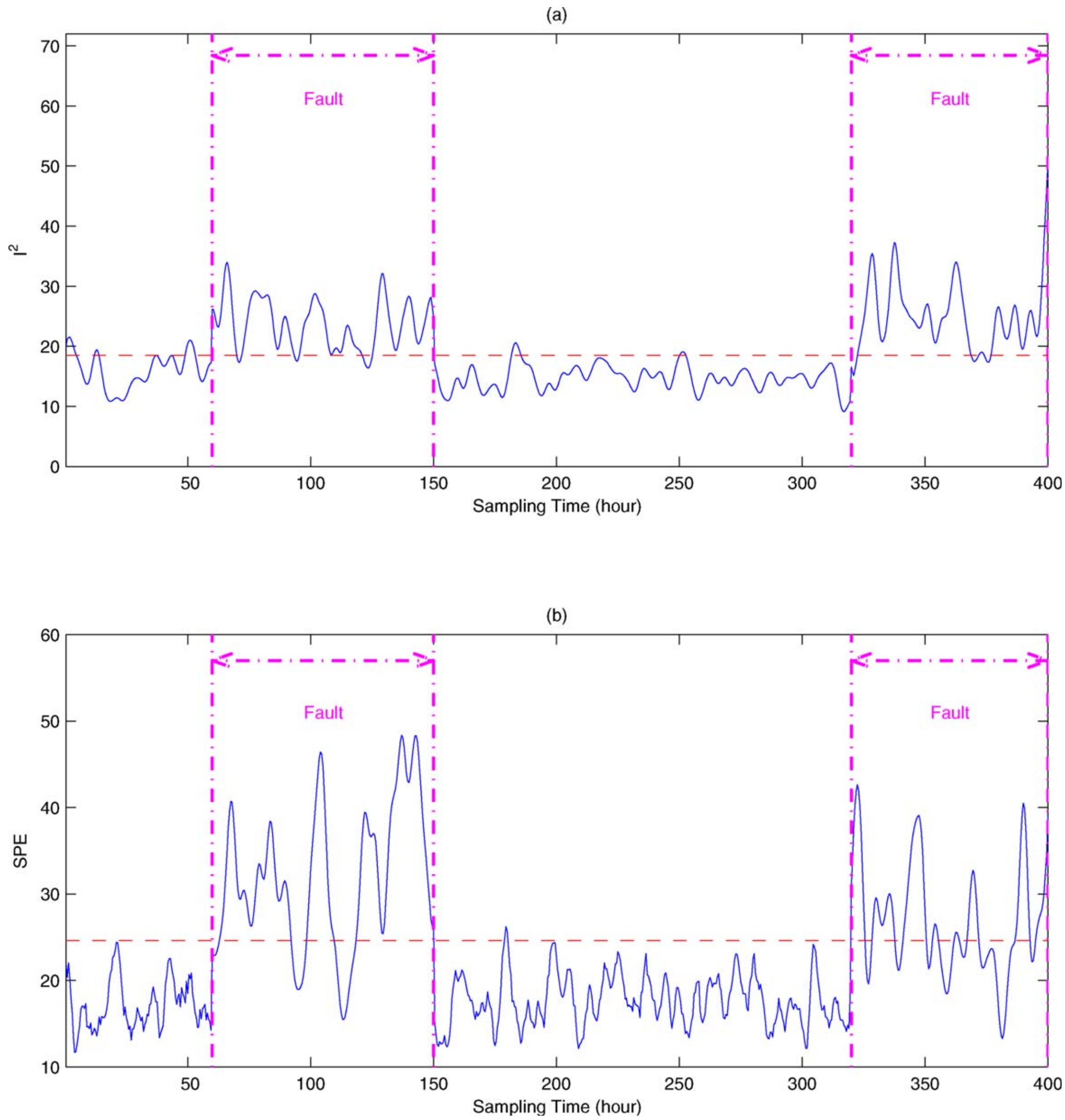

Figure 14. Fault detection results of MICA method in the third test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The second test case also consists of a single process fault, which is a step error in aeration rate with the duration of 140 h. The fault detection results for the MPCA mixture model approach are shown in Figure 9. It is seen that many faulty samples are not alarmed by this approach, although few false alarms are triggered. As shown in Table 3, the false alarm rate of MPCA mixture model approach is as low as 2.99%. However, its corresponding fault detection rate is only 84.70%, which indicates the inferior fault detection capability of the conventional MPCA mixture model method. In Figure 10, the MICA based I^2 index results in substantial percentage of false alarms, although the faulty samples can be detected. The SPE index has improved performance in terms of false alarm rate. Nevertheless, a segment of normal samples toward the end of the batch still trigger false alarms. As opposed to the unsatisfactory fault detection performance of MPCA mixture model and MICA methods, the MICA mixture model based local I^2 and SPE statistics shown in Figure 11 lead to reliable faulty sample detection along with the minimum number of false alarms. The localized MICA models and the adaptively shifted control limits over different operating phases enable the monitoring statistics and their confidence regions better characterizing the normal operating conditions. The quantitative comparison in Table 3 shows that the fault detection and false alarm rates of the proposed approach are 96.26 and 2.5%, whereas the corresponding rates of the conventional MICA method are 92.17 and 24.57%. With the detected faulty samples, the fault diagnoses are further carried out to identify the leading variables with the most dramatic upsets. As shown in Figures 12a, b, the MPCA mixture model approach points to the substrate feed temperature as the major faulty variable while the MICA based contribution plot identifies the variable of substrate concentration as the leading faulty variable with the largest contribution. Apparently, the fault diagnosis results of the two conventional methods are not consistent with the designed faulty scenario as the step error occurs on the aeration rate, which does not have the most significant impact on

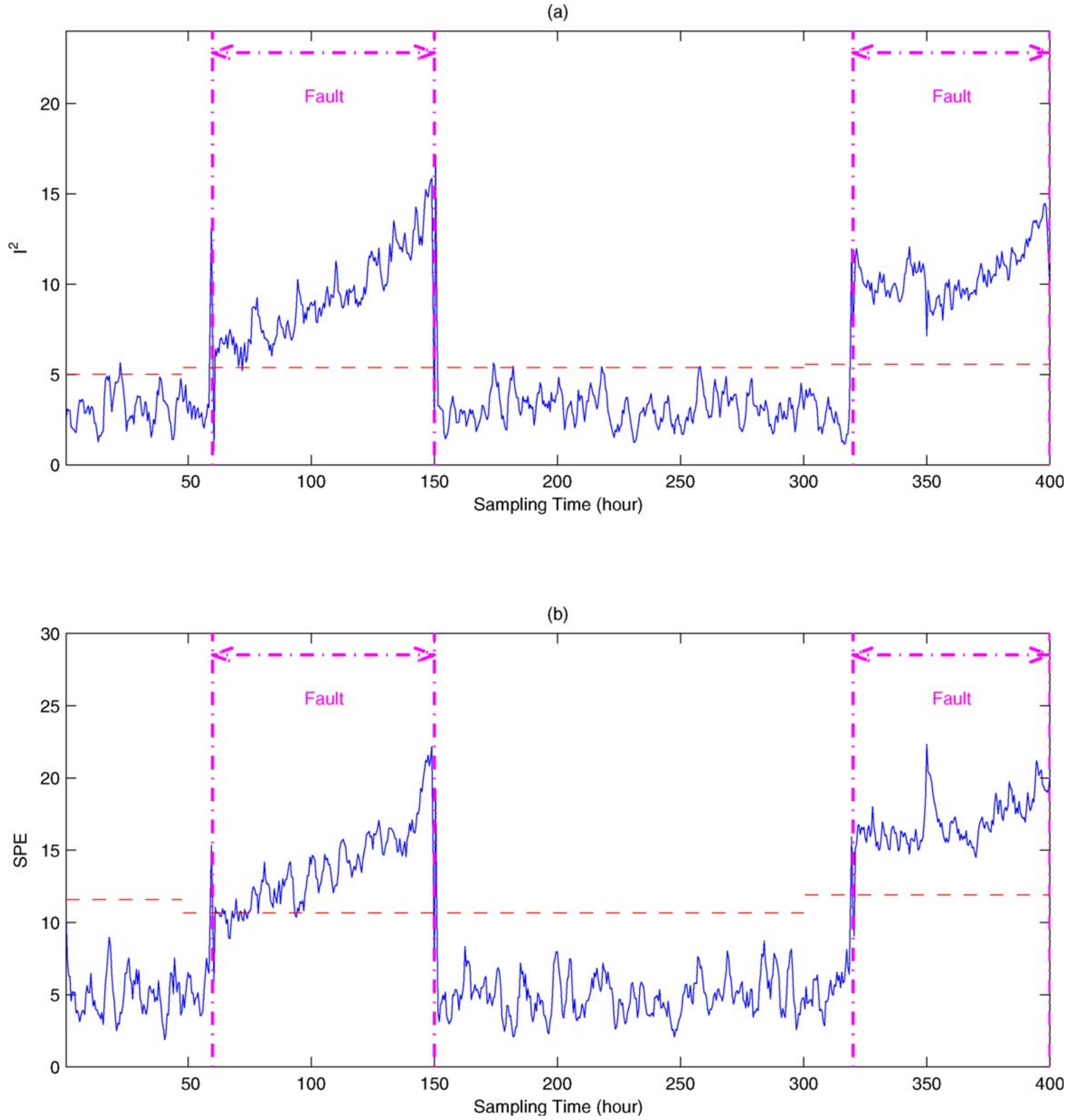

Figure 15. Fault detection results of MICA mixture model based approach in the third test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

either the substrate feed temperature or the substrate concentration. As a comparison, the MICA mixture model and mutual information based contribution plot given in Figure 12b shows that the variable of aeration rate is of the largest contribution and corresponds to the leading variable with the most significant abnormality. Meanwhile, one can easily see that the variable of dissolved oxygen concentration in the fermenter has the second largest contribution, because the abnormal step change in aeration rate can have immediate influence on the oxygen concentration. The identified faulty variables by the proposed method match well the actual fault scenario in the fermentation process.

A more complex faulty scenario is considered in the third test case that involves two types of process faults, a drift error in the substrate feed rate and a step error in the agitator power. The fault detection results of the MPCA mixture model method, MICA method, and the proposed MICA mixture model approach are shown in Figures 13–15, respectively. In this case, the MPCA mixture model approach has low fault detection rate of 86.55% and a high false alarm rate of 16.16%, which is primarily due to the inability of the MPCA models to extract the non-Gaussian hidden features for fault detection. Likewise, quite a number of faulty samples are not detected by the MICA based SPE index, though the I^2 index has a bit better fault detection capability. The average fault detection and false alarm rates of MICA method are 81.14 and 3.6%, respectively. In contrast, both the SPE and I^2 indices of the MICA mixture model approach can correctly detect the vast majority of faulty samples with very few false alarms triggered. The corresponding fault detection and false alarms rates are 95.76 and 1.86%, respectively. Such comparison further demonstrates that the localized MICA models and adaptive control limits can handle the shifting non-Gaussian phases and multiple process faults very well. After the fault detection, the contribution plots for the two identified faulty periods are generated separately. For the first faulty period, the corresponding contribution plots as given in Figures 16a–c show that the MPCA

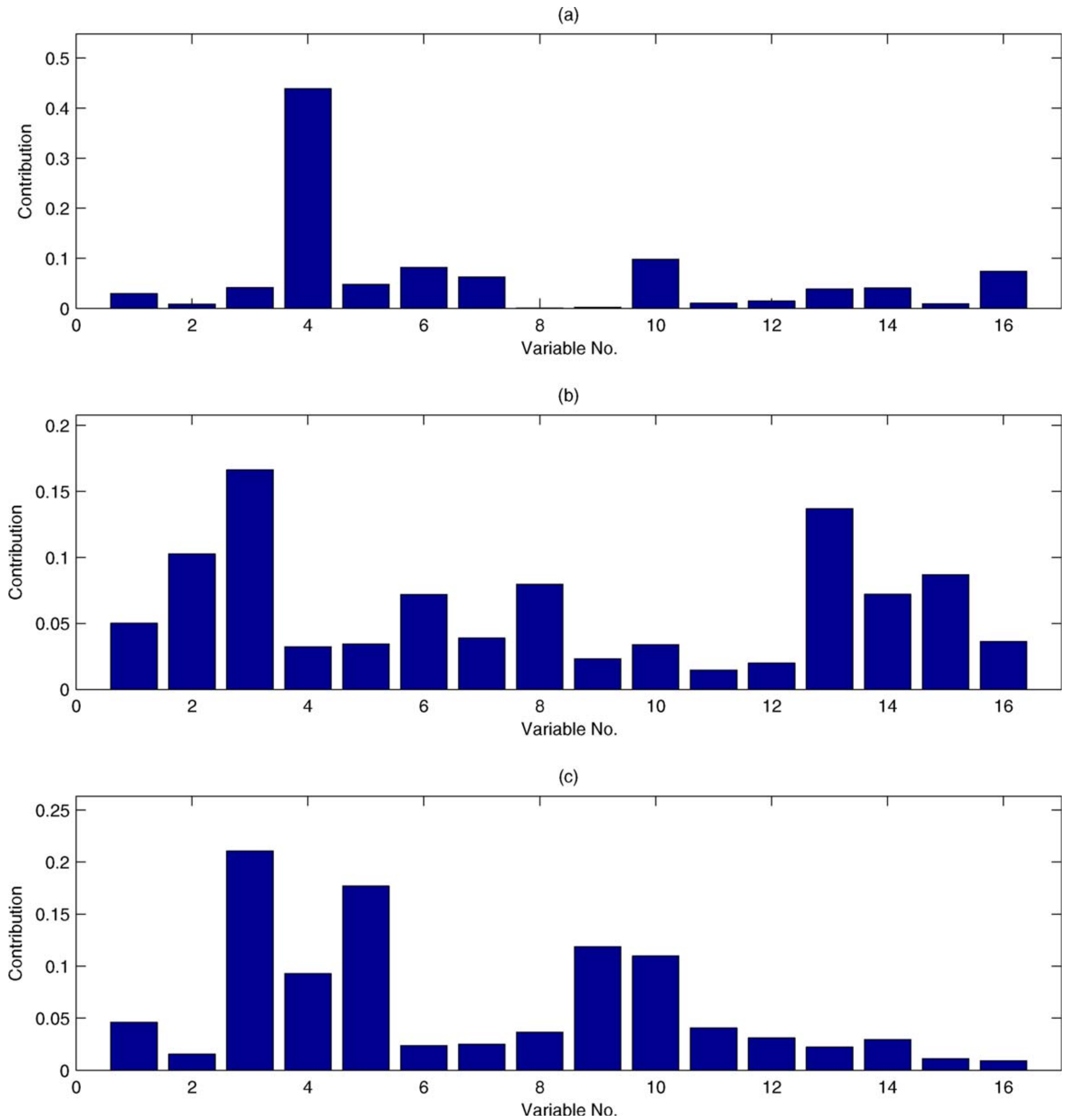

Figure 16. Fault diagnosis results of (a) MPCA mixture model based contribution approach, (b) MICA based contribution approach, and (c) MICA mixture model and mutual information based contribution approach in the first faulty period of the third test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

contribution method incorrectly identifies the substrate feed temperature as the major faulty variable whereas both the MICA and the MICA mixture model diagnosis methods correctly identify the variable of substrate feed rate as the major faulty variable with the largest contribution. However, the MICA based contribution plot also indicates that the variable of generated heat has the second largest contribution, and this observation does not coincide with the process analysis as the drift error in substrate feed rate should cause the upset on the substrate concentration in the fermenter rather than the generated heat. In comparison, the contribution plot of the presented method precisely captures the substrate concentration as the second largest contributor right next to the substrate feed rate. For the second faulty period as shown in Figures 17a–c, the contribution plot of the MPCA mixture model approach still indicates that the highest contribution is incorrectly identified as the substrate feed temperature, whereas the contribution plot of MICA method shows that the variable of pH value has the largest contribution though the fault occurs on agitator power. In contrast, the MICA mixture model and mutual information based contribution plot isolates the agitator power as the leading faulty variable. The above three test cases demonstrate that the MICA mixture model and mutual information based batch process monitoring approach is superior to the conventional MPCA mixture model and MICA based method in terms of significantly enhanced fault detection and diagnosis capability for multiphase non-Gaussian batch processes.

Conclusions

In this article, a novel MICA mixture model and mutual information based approach for fault detection and diagnosis is developed for multiphase batch processes. First, the batch

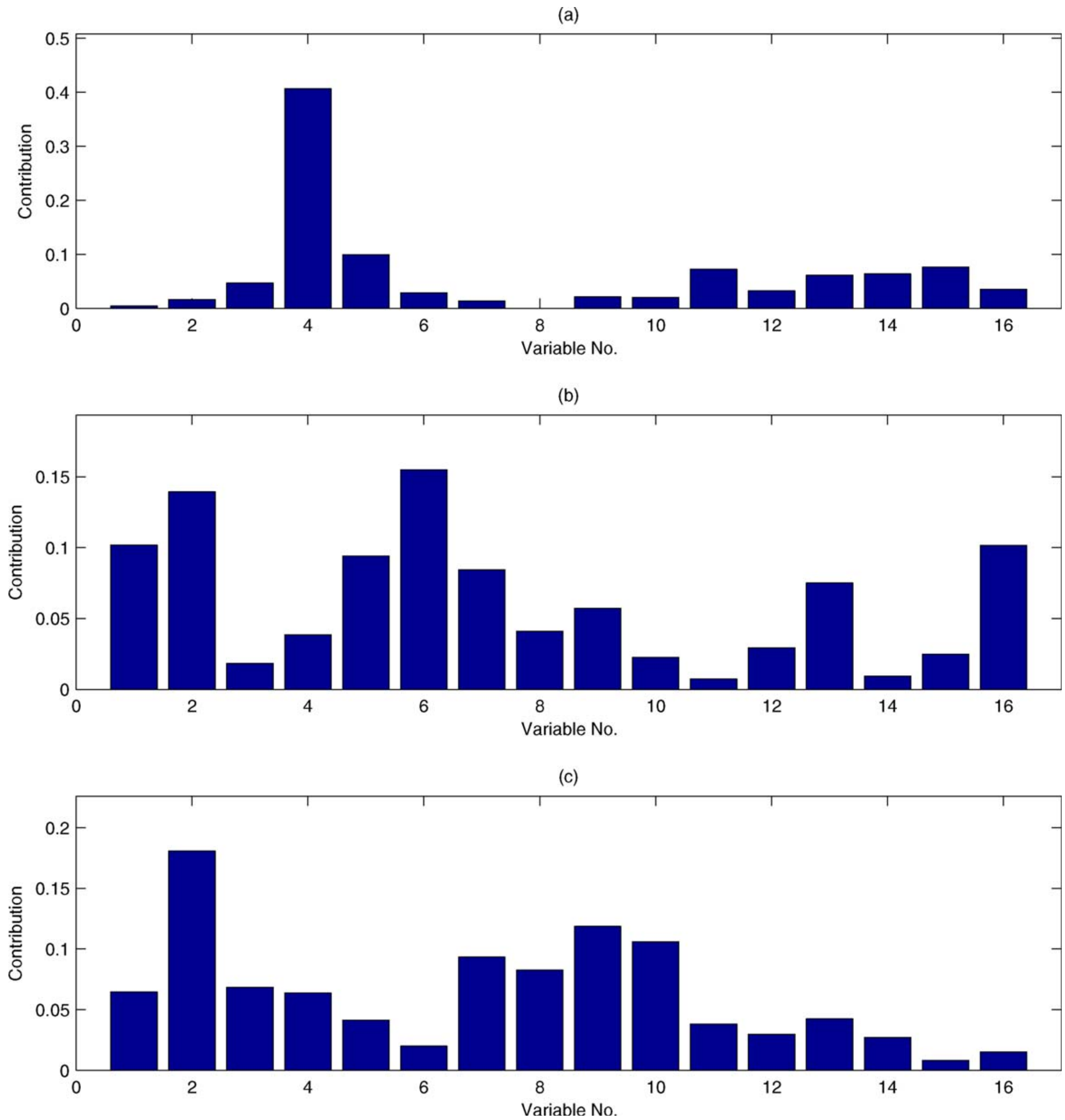

Figure 17. Fault diagnosis results of (a) MPCA mixture model based contribution approach, (b) MICA based contribution approach, and (c) MICA mixture model and mutual information based contribution approach in the second faulty period of the third test case.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

operation with shifting phases and non-Gaussianity within each phase is characterized with the MICA mixture model where the multiple non-Gaussian classes are identified. Then, the posterior probability of the monitored sample is maximized to determine the particular phase the sample belongs to, and, thus, the localized MICA model corresponding to the individual class is built for process fault detection. Furthermore, the detected faulty samples are projected onto the residual subspace, so that a novel mutual information based contribution index is established to measure the statistical dependency between the projection and the original measurement along each process variable. Such contribution index is able to capture the non-Gaussian relationship between individual variables and projected subspace so as to diagnose the major faulty variables with the most significant upsets.

The proposed batch monitoring approach is applied to the fed-batch penicillin fermentation process, and its performance is compared to that of the MPCA mixture model and MICA based monitoring methods. The results show that the proposed approach has superb capability of fault detection and diagnosis for multiphase batch processes with significant within-phase non-Gaussianity. Not only the abnormal events across different phases can be precisely captured by the localized MICA models and adaptively shifted control limits, but also the major faulty variables may be accurately identified by the mutual information based non-Gaussian contribution index. Future work may be focused on extending the idea of non-Gaussian mixture model to phase based adaptive control design of batch or semibatch processes.

Literature Cited

1. Bakshi BR, Locher G, Stephanopoulos G. Analysis of operating data for evaluation, diagnosis and control of batch operations. *J Process Control*. 1994;4:179–194.

2. Lennox B, Montague GA, Hiden HG, Kornfeld G, Goulding PR. Process monitoring of an industrial fed-batch fermentation. *Biotechnol Bioeng.* 2001;74:125–135.
3. Kano M, Hasebe S, Hashimoto I, Ohno H. Statistical process monitoring based on dissimilarity of process data. *AIChE J.* 2002;48: 1231–1240.
4. Chiang LH, Leardi R, Pell RJ, Seasholtz MB. Industrial experiences with multivariate statistical analysis of batch process data. *Chemom Intell Lab Syst.* 2006;81:109–119.
5. Yu J, Qin SJ. Statistical MIMO controller performance monitoring. Part I: data driven covariance benchmark. *J Process Control.* 2008;18:277–296.
6. Yu J, Qin SJ. Statistical MIMO controller performance monitoring. Part II: performance diagnosis. *J Process Control.* 2008;18:297–319.
7. Yu J, Qin SJ. Variance component analysis based fault diagnosis of multi-layer overlay lithography processes. *IIE Trans.* 2009;41:764–775.
8. Yoo C, Lee JM, Vanrolleghem PA, Lee IB. On-line monitoring of batch processes using multiway independent component analysis. *Chemom Intell Lab Syst.* 2004;71:151–163.
9. Yu J. A bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Computer Chem Eng.* 2012;41:134–144.
10. Zorzetto L, Wilson J. Monitoring bioprocesses using hybrid models and an extended kalman filter. *Comput Chem Eng.* 1996;20:689–694.
11. Wold S, Geladi P, Esbensen K, Öhman J. Multi-way principal components-and PLS-analysis. *J Chemom.* 1987;1:41–56.
12. Nomikos P, MacGregor JF. Multi-way partial least squares in monitoring batch processes. *Chemom Intell Lab Syst.* 1995;30:97–108.
13. Kosanovich KA, Dahl KS, Piovoso MJ. Improved process understanding using multiway principal component analysis. *Ind Eng Chem Res.* 1996;35:138–146.
14. Wise BM, Gallagher NB. The process chemometrics approach to process monitoring and fault detection. *J Process Control.* 1996;6: 329–348.
15. Martin EB, Morris AJ. Non-parametric confidence bounds for process performance monitoring charts. *J Process Control.* 1996;6:349–358.
16. Chen J, Liu KC. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem Eng Sci.* 2002;57:63–75.
17. Ündey C, Ertunç S, Çinar A. Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. *Ind Eng Chem Res.* 2003;42:4645–4658.
18. Treasure R, Kruger U, Cooper J. Dynamic multivariate statistical process control using subspace identification. *J Process Control.* 2004;14:279–292.
19. Lee JM, Yoo C, Choi SW, Vanrolleghem PA, Lee IB. Nonlinear process monitoring using kernel principal component analysis. *Chem Eng Sci.* 2004;59:223–234.
20. Acar E, Yener B. Unsupervised multiway data analysis: a literature survey. *IEEE Trans Knowl Data Eng.* 2009;21:6–20.
21. Sun W, Meng Y, Palazoglu A, Zhao J, Zhang H, Zhang J. A method for multiphase batch process monitoring based on auto phase identification. *J Process Control.* 2011;21:627–638.
22. Albazzaz H, Wang XZ. Statistical process control charts for batch operations based on independent component analysis. *Ind Eng Chem Res.* 2004;43:6731–6741.
23. Lee JM, Qin SJ, Lee IB. Fault detection and diagnosis based on modified independent component analysis. *AIChE J.* 2006;52:3501–3514.
24. Zhang Y, Qin SJ. Fault detection of nonlinear processes using multiway kernel independent component analysis. *Ind Eng Chem Res.* 2007;46:7780–7787.
25. Rashid M, Yu J. A new dissimilarity method integrating multidimensional mutual information and independent component analysis for non-Gaussian dynamic process monitoring. *Chemom Intell Lab Syst.* 2012;115:44–58.
26. Rashid M, Yu J. Nonlinear and non-Gaussian dynamic batch process monitoring using a new multiway kernel independent component analysis and multidimensional mutual information based dissimilarity method. *Ind Eng Chem Res.* 2012;51:10910–10920.
27. Yu J. Nonlinear bioprocess monitoring using multiway kernel localized fisher discriminant analysis. *Ind Eng Chem Res.* 2011;50:3390–3402.
28. Yu J. Localized fisher discriminant analysis based complex chemical process monitoring. *AIChE J.* 2011;57:1817–1828.
29. Chiang LH, Russell EL, Braatz RD. Fault Detection and Diagnosis in Industrial Systems. Advanced Textbooks in Control and Signal Processing. London, Great Britain: Springer-Verlag, 2001.
30. Chiang LH, Kotanchek ME, Kordon AK. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Comput Chem Eng.* 2004;28:1389–1401.
31. Mahadevan S, Shah S. Fault detection and diagnosis in process data using one-class support vector machines. *J Process Control.* 2009;19:1627–1639.
32. Yu J, Qin SJ. Multiway Gaussian mixture model based multiphase batch process monitoring. *Ind Eng Chem Res.* 2009;48:8585–8594.
33. Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J.* 2008;54:1811–1829.
34. Yu J. A particle filter driven dynamic Gaussian mixture model approach for complex process monitoring and fault diagnosis. *J Process Control.* 2012;22:778–788.
35. Yu J. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chem Eng Sci.* 2012;68:506–519.
36. Yu J. Multiway discrete hidden Markov model-based approach for dynamic batch process monitoring and fault classification. *AIChE J.* 2012;58:2714–2725.
37. Chen J, Jiang YC. Development of hidden semi-Markov models for diagnosis of multiphase batch operation. *Chem Eng Sci.* 2011;66: 1087–1099.
38. Yao Y, Gao F. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu Rev Control.* 2009;33:172–183.
39. Chen J, Liu J. Mixture principal component analysis models for process monitoring. *Ind Eng Chem Res.* 1999;38:1478–1488.
40. Zhao SJ, Zhang J, Xu YM. Monitoring of processes with multiple operating modes through multiple principal component analysis models. *Ind Eng Chem Res.* 2004;43:7025–7035.
41. Yue H, Qin SJ. Reconstruction based fault identification using a combined index. *Ind Eng Chem Res.* 2001;40:4403–4414.
42. Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw.* 1999;10:626–634.
43. Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. *IEEE Trans Neural Netw.* 2000;13:411–430.
44. Bishop CM. Neural Networks for Pattern Recognition. Oxford, UK: Oxford University Press, 1995.
45. Kassidas A, MacGregor JF, Taylor PA. Synchronization of batch trajectories using dynamic time warping. *AIChE J.* 1998;44:864–875.
46. Lee TW, Lewicki MS, Sejnowski TJ. ICA mixture models for unsupervised classification of non-gaussian classes and automatic context switching in blind signal separation. *IEEE Trans Pattern Anal Mach Intell.* 2000;22:1078–1089.
47. Vejmelka M, Paluš M. Inferring the directionality of coupling with conditional mutual information. *Phys Rev E.* 2008;77:026214.
48. Birol G, Ündey C, Çinar A. A modular simulation package for fed-batch fermentation: penicillin production. *Comput Chem Eng.* 2002;26:1553–1565.

Manuscript received July 8, 2012, and revision received Jan. 19, 2013.