

Research Article

Multimode Process Monitoring Based on Sparse Principal Component Selection and Bayesian Inference-Based Probability

Xiaodong Jiang,¹ Haitao Zhao,¹ and Bo Jin²

¹Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

²Software Engineering Institute, East China Normal University, Shanghai 200062, China

Correspondence should be addressed to Haitao Zhao; ht.zhao.ecust@gmail.com

Received 6 May 2015; Revised 27 July 2015; Accepted 28 July 2015

Academic Editor: Jean J. Loiseau

Copyright © 2015 Xiaodong Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

According to the demand for diversified products, modern industrial processes typically have multiple operating modes. At the same time, variables within the same mode often follow a mixture of Gaussian distributions. In this paper, a novel algorithm based on sparse principal component selection (SPCS) and Bayesian inference-based probability (BIP) is proposed for multimode process monitoring. SPCS can be formulated as a just-in-time regression between all PCs and each sample. SPCS selects PCs according to the nonzero regression coefficients which indicate the compact expression of the sample. This expression is necessarily *discriminative*: amongst all subset of PCs, SPCS selects the PCs which most compactly express the sample and rejects all other possible but less compact expressions. BIP is utilized to compute the posterior probabilities of each monitored sample belonging to the multiple components and derive an integrated global probabilistic index for fault detection of multimode processes. Finally, to verify its superiority, the SPCS-BIP algorithm is applied to the Tennessee Eastman (TE) benchmark process and a continuous stirred-tank reactor (CSTR) process.

1. Introduction

Over the past two decades, with the development of complex chemical processes and the growing demand of plant safety and stable product quality, timely process monitoring is gaining importance. Because large amounts of data can be gathered by the use of distributed control systems (DCSs), multivariate statistical process monitoring (MSPM) algorithms have received great attention. Among these algorithms, principal component analysis (PCA) and partial least squares (PLS) are the most widely used algorithms [1–8]. Both algorithms project high-dimensional data onto lower dimensional subspaces. Process normal and abnormal conditions can be isolated by the use of Hotelling's T^2 or squared predicted error (SPE) [9–13]. Other complementary MSPM algorithms, including independent components analysis (ICA), Fisher discriminant analysis (FDA), and canonical variate analysis (CVA), are used to overcome some limitations in PCA/PLS-based monitoring schemes [14–17]. However, most of MSPM algorithms rely on the assumption

that the system is in a single operating region and that the data follow a Gaussian distribution. In chemical processes, operating condition shifts are often encountered due to the changes of various factors such as feedstock, product specification, set points, and manufacturing strategy. When a process is running under substantially different operating conditions, only a small number of variables actually follow Gaussian distribution [18]. As a result, the multimodality of data distribution might lead to unseemliness for the monitoring of conventional MSPM algorithms. To address these problems, it is necessary to develop new algorithms.

In literature, multiple models can be built to fit each individual operating mode, but these are two essential issues that need to be addressed. One is how to divide the training data into multiple subsets correctly, corresponding to different operating modes. In order to solve this issue, many clustering algorithms are applied. In terms of the traditional approaches, Ge and Song [19] used fuzzy C-means clustering algorithm to separate the training data set according to the unique characteristics of each mode. He et al. [20] applied

the k -nearest-neighbor method. Srinivasan et al. [21, 22] identified the different operating modes by evaluating the Euclidean distances between samples in a constructed data window and then applied dynamic PCA-based similarity measures to cluster the samples. Liu and Chen [23] developed a method using Bayesian classification for selecting multiple regions from a training data set. Zhao et al. [24] presented a multiple principal component analysis (MPCA) algorithm that selects one suitable model to monitor multimode processes. The other issue is how to determine the final results. A proper measurement should be employed to determine which model is the most suitable one for monitoring at the current moment. Ng and Srinivasan [22] exploited the most suitable PCA model through a minimized distance reflecting both the T^2 and SPE values. Zhao et al. [24] close the local PCA model with the minimum SPE value. Natarajan and Srinivasan [21] used the distance between the sample and the center of local models as a criterion. Yu and Qin [25] performed Bayesian inference on the postprobabilities calculated by the Gaussian mixture model (GMM) or the nonlinear kernel GMM. Meanwhile, Ge and coworkers [26, 27] took advantage of Bayesian inference to softly combine the monitoring results computed by local models built by means of probabilistic PCA (PPCA), factor analysis (FA), or subspace algorithms.

To date, the problem of how to correctly divide the training data into multiple subset can successfully be solved by many algorithms mentioned in the previous paragraph. However, there are still some issues that need to be resolved; the most important one is how to select the key principal components (PCs) when using one suitable model for process monitoring. Many algorithms for selecting PCs have been proposed, such as cumulative percent variance (CPV) [28], variance of reconstruction error (VRE) [29], and cross validation (CV) [30]. Generally, most of the classical algorithms just take normal operational observations into account and select the first several PCs with larger variance. While PCs with larger variance of normal data cannot guarantee the capture of the largest variations in fault data online. Jolliffe [31] suggested that the last PCs may be as important as those with large variance. Togkalidou et al. [32] noted that the PCs with larger variance do not always contain much information for prediction. However, this issue is insufficiently discussed in PCA-based process monitoring, and the standard PC selection is still not established.

Fortunately, many researches have been aware of the inherent defects of classical PCA algorithm. A lot of workers tried to seek a subspace spanned by key PCs, which contains the most important information for process monitoring. Peng et al. [33] suggested a new feature selection algorithm, named minimal-redundancy-maximal-relevance criterion (mRMR). It is based on mutual information and selects the features with highest relevance to the target class. Jiang et al. [34] put forward the sensitive principal component analysis for fault detection and diagnosis in chemical processes. They pointed out that PCs selected by PCA algorithm are not always the key PCs for fault detection. Their task was to find the sensitive PCs which have

relationship with fault information. Arbel et al. [35] proposed that the process variables that are preponderant in achieving specific objectives need to be selected.

In this paper, a process monitoring algorithm using multisubspace sparse principal component analysis with the BIP algorithm is put forward. First, variables are divided into different subblocks corresponding to different units or pieces of equipment to reduce the complexity of process analysis. By using BIP algorithm, multimode data in each subblock are divided into multiple subgroups. BIP can compute the posterior probabilities of each monitored sample belonging to the multiple components and derive an integrated global probabilistic index for fault detection of multimode processes. The PCs selected by PCA algorithm with larger variances do not always have relationship with fault information. Sparse principal component selection (SPCS) takes the information of both normal and abnormal observations into account. The algorithm is formulated as a just-in-time form that constructs an elastic net regression between all PCs and each sample. SPCS selects PCs corresponding to the nonzero regression coefficients which indicate the compact expression of the sample. This expression is necessarily *discriminative*: amongst all subset of PCs, SPCS selects the PCs which most compactly express the sample and rejects all other possible but less compact expression. Third, the key PCs are selected by SPCS in each subgroup to solve the problem of fault information loss. It needs to be stressed that the subspace spanned by the key PCs selected is the feature subspace. Finally, in order to verify the superiority of the SPCS-BIP algorithm, it is applied to the Tennessee Eastman (TE) benchmark problem and a continuous stirred-tank reactor (CSTR) process.

2. Preliminaries

2.1. Principal Component Analysis. Principal component analysis is a multivariate statistical analysis which is widely used in chemical process monitoring, fault detection, and so forth [36–38]. Let $\mathbf{x} \in \mathbb{R}^m$ represent an m -dimensional sample vector and $\mathbf{X} \in \mathbb{R}^{N \times m}$ denote a data matrix with zero mean and unit variance, where N is the number of samples and m is the number of variables in the process. From the statistical viewpoint, the PCA algorithm could be obtained by singular value decomposition (SVD) [28, 34]:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \widehat{\mathbf{X}} + \mathbf{E}, \quad (1)$$

where $\mathbf{T} \in \mathbb{R}^{N \times k}$ and $\mathbf{P} \in \mathbb{R}^{m \times k}$ are the score matrix and the loading matrix, respectively. k is the principal components retained number. The loading matrix \mathbf{P} can be obtained by eigenvalue decomposition on the covariance matrix $\text{cov}(\mathbf{x})$ as follows:

$$\text{cov}(\mathbf{x}) \approx \frac{1}{N-1} \mathbf{X}^T \mathbf{X} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T, \quad (2)$$

where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ denotes the eigenvalue matrix and $\mathbf{P} = [\widehat{\mathbf{P}} \widetilde{\mathbf{P}}]$ contains the loading matrices of component subspace and residual subspace, respectively.

2.2. *Construction of Finite Gaussian Mixture Model Based on EM.* For the process running at multiple operating condition, owing to the mean shifts or covariance changes, the assumption of multivariate Gaussian distribution becomes invalid [21, 22]. In this situation, the local Gaussian distribution is still appropriate to characterize each subset of measurement data from the same operating conditions. Therefore, the finite Gaussian mixture model is prime suited to represent the data sources driven by different operating modes [13, 24, 25].

To construct a FGMM, given a set of training samples $\mathbf{X} \in \mathbb{R}^{N \times m}$, the log-likelihood function can be expressed as

$$\log L(\mathbf{X}, \Theta) = \sum_{j=1}^N \log \left(\sum_{i=1}^L \omega_i g(x_j | \theta_i) \right) \quad (3)$$

and the parameter estimation problem is formulated as

$$\widehat{\Theta} = \arg \max_{\Theta} (\log L(\mathbf{X}, \Theta)), \quad (4)$$

where $\Theta = \{\{\omega_i, \mu_i, \Sigma_i\}, \dots, \{\omega_L, \mu_L, \Sigma_L\}\}$, ω_i ($1 \leq i \leq L$) is the prior probabilities, and L is the number of Gaussian components included in FGMM. μ_i is the mean vector and Σ_i is the covariance matrix.

There are a lot of learning algorithms, such as maximum likelihood estimation (MLE), EM, and the F-J algorithm, that have been put forward for mixture model estimation [39, 40]. As a more tractable numerical strategy, the EM algorithm has been well used in practice to estimate the maximum likelihood distribution parameter [39]. EM algorithm is implemented iteratively by means of repeating the expectation step (E-step) and maximization step (M-step) to calculate the posterior probabilities and then the corresponding distribution parameters until a convergence criterion of the log-likelihood function is satisfied. Given the training data \mathbf{X} and an initial estimate $\Theta^{(0)} = \{\{\omega_1^{(0)}, \mu_1^{(0)}, \Sigma_1^{(0)}\}, \dots, \{\omega_L^{(0)}, \mu_L^{(0)}, \Sigma_L^{(0)}\}\}$, the iterative E-step and M-step are expressed as follows:

(i) E-step:

$$P^{(s)}(C_l | \mathbf{x}_j) = \frac{\omega_l^{(s)} g(\mathbf{x}_j | \mu_l^{(s)}, \Sigma_l^{(s)})}{\sum_{i=1}^L \omega_i^{(s)} g(\mathbf{x}_j | \mu_i^{(s)}, \Sigma_i^{(s)})}, \quad (5)$$

where $P^{(s)}(C_l | \mathbf{x}_j)$ denotes the posterior probability of the j th training sample within the l th Gaussian component at the s th iteration;

(ii) M-step:

$$\begin{aligned} \mu_l^{(s+1)} &= \frac{\sum_{j=1}^N P^{(s)}(C_l | \mathbf{x}_j) \mathbf{x}_j}{\sum_{j=1}^N P^{(s)}(C_l | \mathbf{x}_j)}, \\ \Sigma_l^{(s+1)} &= \frac{\sum_{j=1}^N P^{(s)}(C_l | \mathbf{x}_j) (\mathbf{x}_j - \mu_l^{(s+1)}) (\mathbf{x}_j - \mu_l^{(s+1)})^T}{\sum_{j=1}^N P^{(s)}(C_l | \mathbf{x}_j)}, \\ \omega_l^{(s+1)} &= \frac{\sum_{j=1}^N P^{(s)}(C_l | \mathbf{x}_j)}{N}, \end{aligned} \quad (6)$$

where $\mu_l^{(s+1)}$, $\Sigma_l^{(s+1)}$, and $\omega_l^{(s+1)}$ are the mean, covariance, and prior probability of the l th Gaussian component at the $(s+1)$ th iteration, respectively.

3. Fault Detection with Sparse Principal Component Selection and Bayesian Inference-Based Probability

In this section, the idea of SPCS-BIP algorithm for multimode process monitoring is demonstrated in detail. We first introduce the Bayesian inference-based probability which can derive the confidence boundary around the normal operating regions for process monitoring and fault detection. Then, the sparse principal component selection was introduced for selecting the key PCs related with fault information. Finally, the steps of this algorithm were given.

3.1. *Bayesian Inference-Based Probability.* In the previous section, the FGMM has been constructed, and it is essential to further derive the confidence boundary around the normal operating regions for process monitoring and fault detection. Due to the multimodality of mixture distribution, it is really difficult to capture the analytical boundary of the density function $p(x | \Theta)$ in a certain confidence level.

In the proposed monitoring approach, given an arbitrary monitored sample x_t belonging to each Gaussian component, Bayesian inference strategy is used to calculate the posterior probability as follows:

$$\begin{aligned} P(x_t \in C_l) &= P(C_l | x_t) = \frac{P(C_l) \cdot p(x_t | C_l)}{p(x_t)} \\ &= \frac{P(C_l) \cdot p(x_t | C_l)}{\sum_{i=1}^L P(C_i) \cdot p(x_t | C_i)} \end{aligned} \quad (7)$$

which can also be formulated as

$$P(x_t \in C_l) = \frac{\omega_l g(x_t | \mu_l, \Sigma_l)}{\sum_{i=1}^L \omega_i g(x_t | \mu_i, \Sigma_i)} \quad (l = 1, 2, \dots, L). \quad (8)$$

Given that each component C_l follows a unimodal Gaussian distribution, the squared Mahalanobis distance of x_t from the center of C_l follows χ^2 distribution, provided that x_t belongs to C_l ,

$$D((x_t, C_l) | x_t \in C_l) = (x_t - \mu_l)^T \Sigma_l^{-1} (x_t - \mu_l) : \chi_m^2. \quad (9)$$

Under the assumption that $x_t \in C_l$ and χ_m^2 has m degree of freedom, $D((x_t, C_l) | x_t \in C_l)$ denotes the squared Mahalanobis distance between x_t and the mean center of C_l . Owing to colinearity, Σ_l is usually ill-conditioned, and the following regularized Mahalanobis distance is utilized instead to avoid too wide confidence regions:

$$\begin{aligned} D_r((x_t, C_l) | x_t \in C_l) \\ = (x_t - \mu_l)^T (\Sigma_l + \varepsilon I)^{-1} (x_t - \mu_l), \end{aligned} \quad (10)$$

where the function of ε is to remove the ill condition of covariance matrix Σ_l by adding a positive number to all the diagonal entries.

For the monitored sample x_t , a local Mahalanobis distance-based probability index relative to each Gaussian component C_l can be defined as

$$P_L^{(l)}(x_t) = \Pr \{D((x, C_l) | x \in C_k) \leq D((x_t, C_l) | x_t \in C_l)\} \quad (11)$$

or

$$P_L^{(l)}(x_t) = \Pr \{D_r((x, C_l) | x \in C_l) \leq D_r((x_t, C_l) | x_t \in C_l)\}. \quad (12)$$

Given the appropriate degree of freedom, $P_L^{(l)}(x_t)$ can be computed by integrating the χ^2 probability density function. Under a given confidence level, this index has the function of indicating whether the monitored sample is normal or abnormal provided that it belongs to the corresponding Gaussian component. A global BIP index is proposed to combine the local probability metrics across all the Gaussian clusters because the random characteristic of each monitored sample may come from multiple Gaussian components with the corresponding posterior probabilities. The formulation of BIP index for the monitored sample x_t is given by

$$\text{BIP} = \sum_{l=1}^L P(C_l | x_t) P_L^{(l)}(x_t), \quad (13)$$

where the posterior probability $P(C_l | x_t)$ is used to incorporate the contribution of each local Gaussian component to the overall probabilistic index. As $0 \leq P_L^{(l)}(x_t) \leq 1$, we have

$$0 \leq \text{BIP} \leq \sum_{l=1}^L P(C_l | x_t) = 1. \quad (14)$$

Under the preset confidence level $(1-\alpha) 100\%$, the process is determined within normal operation if

$$\text{BIP} \leq 1 - \alpha. \quad (15)$$

Otherwise, the process operation is treated out of control.

3.2. Sparse Principal Component Selection. Sparse representation has proven to be an extremely powerful tool for acquiring, representing, and compressing high-dimensional data [41–43]. This success is mainly because of the fact that the important reconstruction information of data such as process data and time series data has naturally sparse representations with respect to fixed bases, or concatenations of such bases. Qiao et al. [44] proposed that the graphs constructed by the L_1 -norm have the advantage of greater robustness to data noise, automatic sparsity, and adaptive neighborhood for individual datum. What is more another important advantage is that sparse representation has the potential discriminative ability since most nonzero elements

are located on the samples in the same class as the represented sample.

Given the training sample $\mathbf{L} = [l_1, l_2, \dots, l_N] \in \mathbb{R}^{N \times m}$, a test sample $\mathbf{H} \in \mathbb{R}^m$, the solution to the sparse representation problem can be obtained by solving the following ℓ^1 -minimization problem:

$$\begin{aligned} (\ell^1) : \hat{\beta} &= \arg \min \|\beta\| \\ \text{subject to } \mathbf{L}\beta &= \mathbf{H} \\ \text{Card}(\beta) &\leq k, \end{aligned} \quad (16)$$

where $\hat{\beta} = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_N]$ are the sparse representation coefficients and $\text{Card}(\beta)$ denotes the number of nonzero elements of β . From the perspective of statistics, formula (16) can be named the Lasso criterion. Lasso is a penalized least squares algorithm which was originally by quadratic programming imposing a constraint on the L_1 norm of the regression coefficients. Thus, the Lasso estimates $\hat{\beta}$ are obtained by minimizing the Lasso criterion:

$$\begin{aligned} \hat{\beta} &= \arg \min_{\beta} \left\| \mathbf{H} - \sum_{i=1}^N \mathbf{L}_i \beta_i \right\|^2 + \lambda \sum_{i=1}^N |\beta_i| \\ \text{subject to } \text{Card}(\beta) &\leq k, \end{aligned} \quad (17)$$

where λ is nonnegative. However, only using the L_1 -norm penalty in Lasso has its limitation. Zou et al. [45] proposed that if there is a group of variables among which the pairwise correlations are very high, lasso tends to select any variable from the group and does not consider which one is selected. Fortunately, elastic net was put forward by Zou et al. which overcomes the limitation of only using the L_1 -norm penalty. It is known that combining the L_1 -norm and L_2 -norm penalty can result in grouping effectiveness in regression and thus enhance the prediction accuracy. For any nonnegative λ_1 and λ_2 , the elastic net estimates $\hat{\beta}_{\text{en}}$ are given by

$$\begin{aligned} \hat{\beta}_{\text{en}} &= \arg \min_{\beta} \left\| \mathbf{H} - \sum_{i=1}^N \mathbf{L}_i \beta_i \right\|^2 + \lambda_1 \sum_{i=1}^N |\beta_i| \\ &\quad + \lambda_2 \sum_{i=1}^N |\beta_i|^2 \end{aligned} \quad (18)$$

$$\text{subject to } \text{Card}(\beta) \leq k.$$

In brief, it is expected that the elastic net is used to group a set of sparse coefficients to construct the sparse alignment matrices, in which the sparse representation information or the potential discriminative information is encoded to enhance the discriminative ability in an unsupervised manner.

3.3. Fault Detection with SPCS and BIP. The key problem for monitoring the multimode process is to select a suitable model and choose the subspace spanned by key PCs. In the Introduction, we had put forward the fact that the subspace

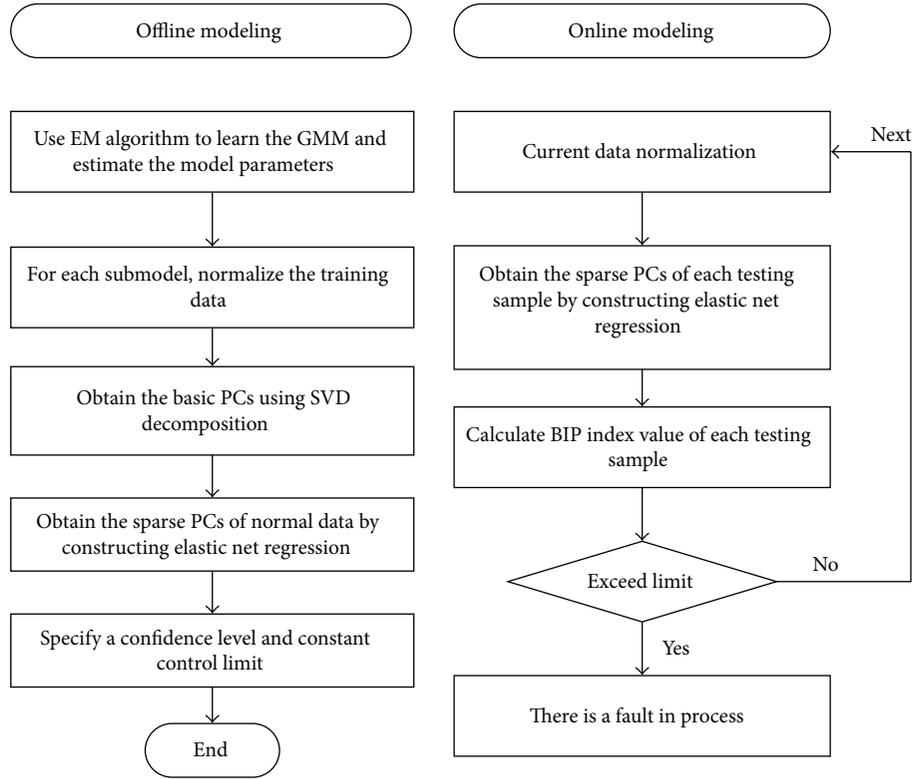


FIGURE 1: The steps of SPCS-BIP algorithm for process monitoring.

spanned by the first several PCs with largest explained variance does not always have fault information.

In the following part, a novel multimode process monitoring approach based on SPCS and BIP is proposed. This approach is in a just-in-time form. For each sample, an elastic net regression between all PCs and the sample is constructed and solved. The PCs which have nonzero regression coefficients are retained, while other PCs are rejected. That means, for each sample, we can pick out the most discriminative bases and the others are set to zero. Its concrete calculating steps are summarized in Figure 1.

Offline Modeling

- (1) Collect a set of historical training data under all possible operating conditions.
- (2) Use the EM algorithm to learn the Gaussian mixture model and estimate the model parameter set $\Theta = \{\mu_1, \Sigma_1, \omega_1, \dots, \mu_k, \Sigma_k, \omega_k\}$ based on the iterative steps.
- (3) For each submodel, get a normal operational observation set $\mathbf{X} = \mathbb{R}^{N \times m}$, where N is the number of samples and m is the number of variables. This set is denoted as the training set for threshold determining. A testing set $\mathbf{Y} \in \mathbb{R}^{S \times m}$ with both normal and abnormal observations is given for testing.
- (4) Normalize the training data through the mean value and variance of each variable.

(5) Obtain all principal components using SVD decomposition. The training data \mathbf{X} is reconstructed by $\mathbf{X} = \sum_{i=1}^m \mathbf{t}_i \mathbf{p}_i^T$, where \mathbf{t}_i is the score vector and \mathbf{p}_i is the loading vector.

(6) For training sample \mathbf{x}_j ($j = 1, 2, \dots, N$), construct an elastic net regression between each observation value of training data and loading vector \mathbf{P} made of PCs in step (5), according to $\hat{\beta}_j = \arg \min_{\beta} \|\mathbf{x}_j - \sum_{i=1}^m \mathbf{p}_i \beta_i\|^2 + \lambda_1 \sum_{i=1}^m |\beta_i| + \lambda_2 \sum_{i=1}^m \|\beta\|^2$ subject to $\text{Card}(\beta) \leq k$.

(7) Corresponding to the nonzero representation coefficients $\{\hat{\beta}_{j_1}, \hat{\beta}_{j_2}, \dots, \hat{\beta}_{j_k}\}$, construct a new loading vector $\hat{\mathbf{P}}_j = [\mathbf{p}_{j_1}, \mathbf{p}_{j_2}, \dots, \mathbf{p}_{j_k}]$.

(8) Specify a confidence $(1-\alpha) 100\%$ and constant control limit $1-\alpha$.

Online Monitoring

- (1) Normalize the current time point data by using mean values and variance of the training data.
- (2) Obtain the loading vector \mathbf{P} from offline modeling.
- (3) When a test sample $\mathbf{y}_j \in \mathbb{R}^m$ ($j = 1, 2, \dots, S$) is coming, construct an elastic net regression between the sample and loading vector \mathbf{P} made of PCs in step (2), according to $\hat{\beta} = \arg \min_{\beta} \|\mathbf{y}_j - \sum_{i=1}^m \mathbf{p}_i \beta_i\|^2 + \lambda_1 \sum_{i=1}^m |\beta_i| + \lambda_2 \sum_{i=1}^m \|\beta\|^2$ subject to $\text{Card}(\beta) \leq k$.

- (4) Corresponding to the nonzero representation coefficients $\{\hat{\beta}_{l_1}, \hat{\beta}_{l_2}, \dots, \hat{\beta}_{l_k}\}$, construct a new loading vector $\widehat{\mathbf{P}}_j = [\mathbf{p}_{l_1}, \mathbf{p}_{l_2}, \dots, \mathbf{p}_{l_k}]$.
- (5) Generate the BIP control chart with the calculated BIP index values for all the monitored samples. If the BIP index of a test sample is lower than the control limit, which means the sample is normal, go to step (1). Else, there is a fault in the process.

4. Case Studies on the TE and CSTR Process

In this case study, the TE benchmark and CSTR process are introduced to verify the effectiveness of the SPCS-BIP algorithm. PCA-GMM is the classic algorithm for multimode processing monitoring. And the fault detection index (FDI) is similar to Bayesian inference probability (BIP). So here, a comparison was made between SPCS-BIP and PCA-GMM. In addition, to verify the improvements of SPCS algorithm, which can select sparse PCs, a comparison was performed between the SPCS-BIP algorithm and the MPPCA algorithm.

4.1. Tennessee Eastman Process. As a well-known benchmark process, the Tennessee Eastman process, which was presented by Downs and Vogel, has been widely applied to evaluate and compare the efficiency of process monitoring techniques [46, 47]. The schematic diagram of the process is illustrated in Figure 2. This process consists of five major unit operations: a reactor, a product condenser, a vapor-liquid separator, a recycle compressor, and a product stripper. In addition, there are six modes of process operation, as listed in Table 1. The variables can be divided into three categories: composition variables, continuous process variables, and manipulated variables. In our study, only modes 1 and 3 were simulated through the Simulink programs developed on the basis of the decentralized control strategy designed by Ricker [48]. The Simulink programs can be downloaded from <http://depts.washington.edu/control/LARRY/TE/download.html>. The 31 selected monitoring variables contained 9 manipulated variables and 22 continuous process variables. Thus, these variables were divided into five subblocks according to five units. However, given that only two variables were allocated to each of the compressor unit and the condenser unit, there were four variables assigned to the other three related subblocks. As a result, the total of 31 variables was divided into three subblocks.

There are 20 faults in the multimode TE process, which are listed in Table 2. Among these faults, the root causes of the faults 16–20 are unknown [46, 47]. What is more, to simplify interpretation, the amplitudes of faults 3, 9, and 15 are so small. It is difficult to detect, so only the remaining 12 faults were considered in this study. In the modeling stage, 2000 normal samples, which include 1000 mode 1 samples and 1000 mode 3 samples, were collected as the training data set. In the testing stage, 1000 samples of mode 1 were tested first, and then the process switches to mode 3. As a result, the test data set consists of 1000 samples of mode 1 and 1000 samples of mode 3. And faults occurred from the 1200th sample. A set of 20 faults in multimode TE process, which are listed

TABLE 1: Six process operation modes of TE process.

Mode	G/H mass ratio	Production rate
1	50/50	7038 kg/h G and 7038 kg/h H
2	10/90	1048 kg/h G and 12669 kg/h H
3	90/10	10000 kg/h G and 11111 kg/h H
4	50/50	Maximum
5	10/90	Maximum
6	90/10	Maximum

in Table 2, are simulated and the corresponding process data are collected for testing. The following simulations are run in MATLAB 8.3.0 (2014a) environment. Here, two indicators, which are FR (FR) and MR (MR), are often introduced to measure the result of process monitoring. FR is the rate of normal data classified as fault data. MR is the rate of fault data classified as normal rate.

In the MPPCA algorithm and PCA-GMM algorithm, when the variance contribution was selected as 85%, the dimension of feature space in MPPCA and the number of PCs in PCA-GMM were each selected as 18. In order to compare the monitoring performances of these algorithms in the same situation, the selected sparse PCs of each mode in SPCS-BIP were selected as 18. The 99% control limit was assigned to all three algorithms.

First, Figure 3 shows that the different submodes can be successfully divided by the EM algorithm used in this paper. And by using other algorithms, the modes also can be divided correctly. In other words, how to correctly divide the training data into multiple subset is not a problem by many related algorithms.

The normal process was tested by different algorithm, and the results are shown in Figure 4. In this figure, it is hard to figure out which algorithm's FR is lower. In the figure, most samples of each algorithm are lower than the control limit. And by calculation, the FR of these algorithms are 0.333%, 0.08%, 0.25%, and 1.08%, respectively, corresponding to Figures 4(a), 4(b), 4(c), and 4(d). The monitoring performances of these three algorithms suggest that the FR are acceptable. Next, the data sets of 12 faults in mode 3 were tested, and the MR of these three algorithms are listed in Table 3, with the smallest MR shown in bold.

From Table 3, we observe that the monitoring performance of SPCS-BIP is the best, compared to the MPPCA and PCA-GMM algorithms for all 12 faults. Here, we take the further analysis. In comparison with MPPCA and PCA-GMM algorithms, the SPCS-BIP algorithm can exactly divide the process data into subgroups corresponding to different modes by using the E-M algorithm, and, in each submode, SPCS-BIP can select the most important PCs that have most relation with the fault. Due to the fact that the subspace spanned by the PCs was monitored by BIP, most of the PCs are related to the main process of chemical industrial process, and only little PCs are related to the fault process. SPCS are discriminative by constructing an elastic net regression between all PCs and each sample. So, in Table 3, we observe

TABLE 2: Process faults for the multimode TE process.

Faults number	Disturbance state	Type
IDV(1)	A/C feed ratio, B composition constant (Stream 4)	Step
IDV(2)	B composition, A/C ratio constant (Stream 4)	Step
IDV(3)	D feed temperature (Stream 2)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss (Stream 1)	Step
IDV(7)	C header pressure loss reduced availability (Stream 4)	Step
IDV(8)	A, B, and C feed composition (Stream 4)	Random variation
IDV(9)	D feed temperature (Stream 2)	Random variation
IDV(10)	C feed temperature (Stream 4)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	Unknown	Unknown
IDV(17)	Unknown	Unknown
IDV(18)	Unknown	Unknown
IDV(19)	Unknown	Unknown
IDV(20)	Unknown	Unknown

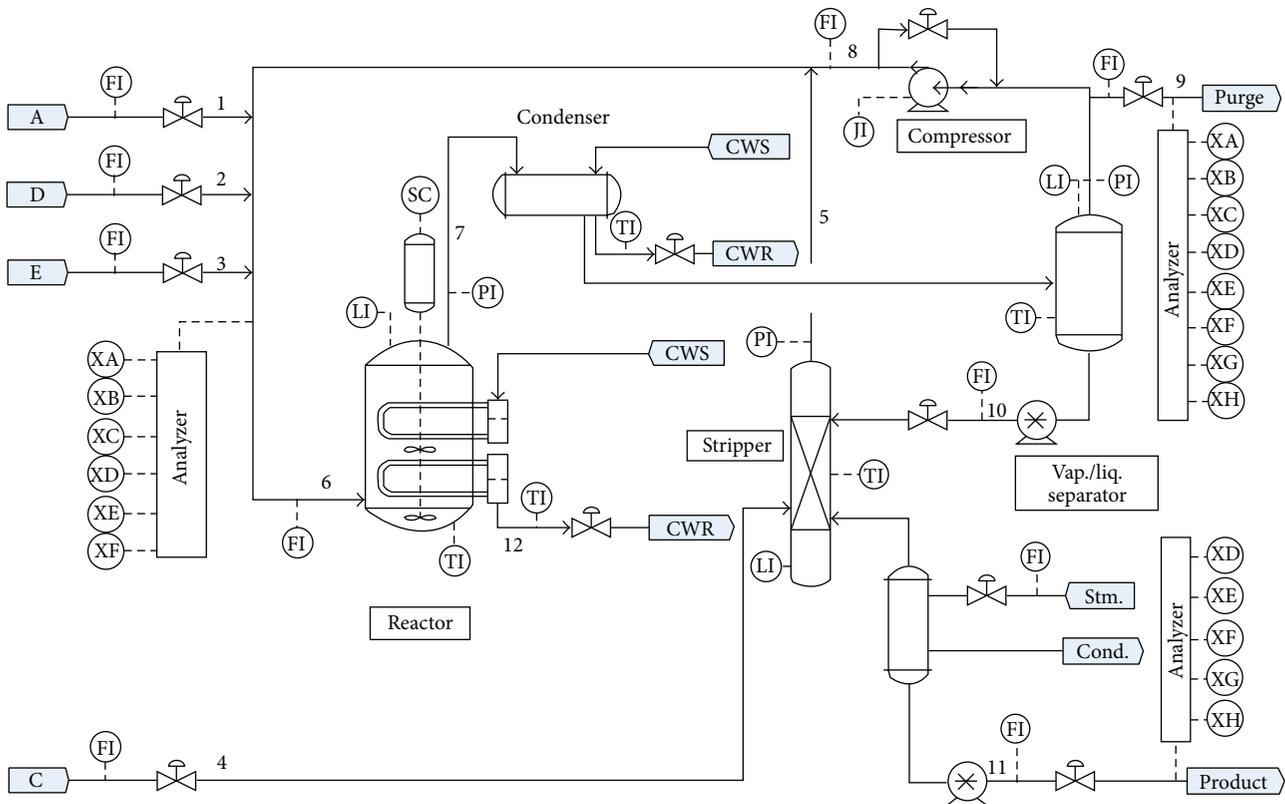


FIGURE 2: Control scheme for the TE process.

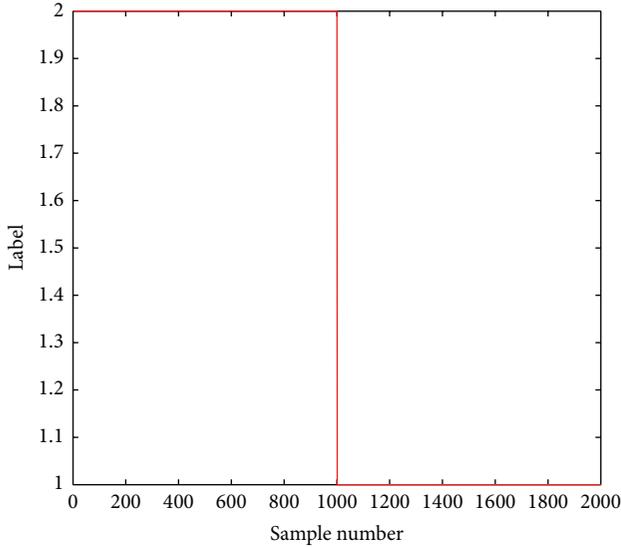


FIGURE 3: Different modes of the training data.

TABLE 3: Missed detection rates (%) of 12 faults.

Faults number	MPPCA- T^2	MPPCA-SPE	PCA-GMM	SPCS-BIP
1	0.75	0.25	1	0.125
2	3.5	8.5	5.5	1.875
4	0	0	0	0
5	0.375	0	1.125	0
6	0	0	0	0
7	0	0	0	0
8	3.375	3.875	4.25	2.500
10	82.375	16	89.875	7.875
11	2.25	8.5	4.5	1.125
12	1.375	1.625	1.5	0.750
13	16.125	26.5	1.9625	11.625
14	0	3.5	0.125	0

that the results of SPCS-BIP are better than the results of MPPCA and PCA-GMM.

Figure 5 shows the monitoring performances of fault 10. It is easy to see that the FDI of MPPCA- T^2 and the BIP of PCA-GMM cannot detect the fault effectively in Figures 4(a) and 4(c). In the figure, more than half of the fault samples were regarded as the normal samples, while, compared to the performances of MPPCA- T^2 and PCA-GMM, the FDI of MPPCA-SPE shows some improvements. However, the monitoring performance of MPPCA-SPE does not match the performance of SPCS-BIP. We can find this point both in Figure 4 and Table 3

4.2. CSTR. This study simulated the CSTR process described by Yoon and MacGregor [49]. The diagram of the process is presented in Figure 6. Due to the fact that the CSTR process consisted of only one operating unit, the number of subblocks was selected as 1.

TABLE 4: Missed detection rates (%) of two faults.

Faults number	MPPCA- T^2	MPPCA-SPE	PCA-GMM	SPCS-BIP
1	90	100	98.8	65.6
2	28.8	99.6	31.8	28

In the modeling stage, 1000 samples which include 500 mode 1 samples and 500 mode 2 samples were collected as the training data set. In the testing stage, 1000 samples of mode 2 were tested, and two faults were introduced to the process as follows.

Case 1. A step of 1K was added in the cooling water temperature T_C from the 500th sample.

Case 2. A 2 kmol/(m³ · min) step was added in the inlet solute concentration C_{AA} from the 500th sample.

In the MMPCA algorithm, when the variance contribution was selected as 85%, the dimension of feature space in MPPCA is 10. So, in order to compare the monitoring performances of these algorithms in the same situation, the number of PCs in PCA-GMM and the selected sparse PCs in SPCA-BIP were both selected as 10. The 99% control limit was assigned to all three algorithms.

The same as TEP, the FR of these algorithms are 0.4%, 0, 1.2%, and 1%, respectively. In an industry process, FR lower than 0.05 is acceptable [28].

The data sets of two faults in mode 2 were tested, and the MR were listed in Table 4. In the table, the smallest missed detection rates are shown in bold.

As shown in Table 4, the SPCS-BIP algorithm has shown the best performance for these two faults compared with other algorithms listed in the table. It is obvious that neither MPPCA nor PCA-GMM algorithms can detect the fault because their missed detection rates were high. In those four algorithms, only the SPCS-BIP was based on the selection PCs, so the improvements in the proposed sparse principal components selection can be demonstrated through the better monitoring performance of the SPCS-BIP algorithm.

Fault 1 is a bias in cooling water temperature T_C . Due to the control loop in the CSTR process, these would be a bias in outlet temperature T , and then the cooling water flow rate F_C would increase. In Figure 7, both the MMPCA and PCA-GMM algorithms could not detect fault 1 effectively according to the performances of those shown in Figures 7(a), 7(b), and 7(c). In Figure 7(d), it is obvious that the monitoring performance of SPCS-BIP is much better than the others. The reason is that the correct classification for each subgroup by using E-M algorithm and the PCs selected by SPCS are discriminative and could construct the subspace that contains the important fault information for abnormal data.

Fault 2 is a bias in inlet solute concentration C_{AA} . Then, due to the control loop in the CSTR process, there would be biases in outlet concentration C and outlet temperature T . According to the performances shown in Figure 8(b), the FDI of MPPCA-SPE could not detect fault 2. Compared to the MPPCA-SPE, the FDI of MPPCA- T^2 showed some

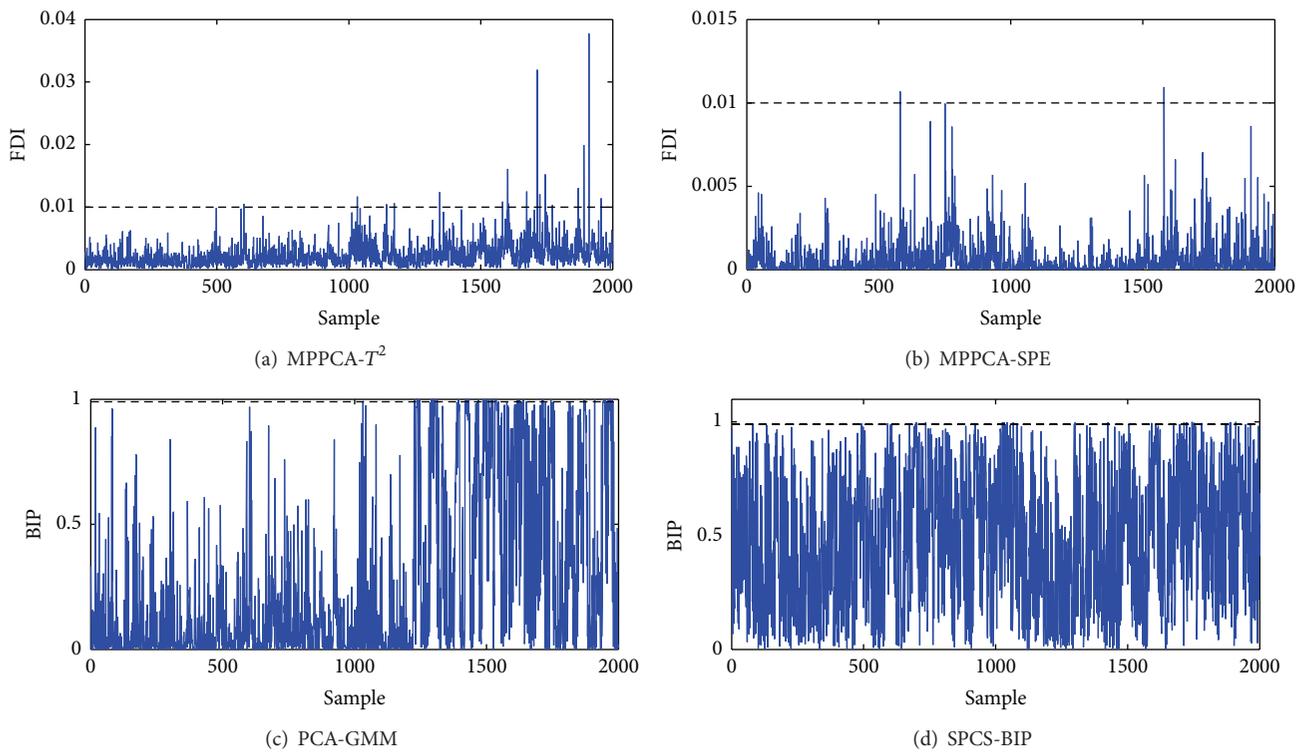


FIGURE 4: Monitoring performance of the normal process.

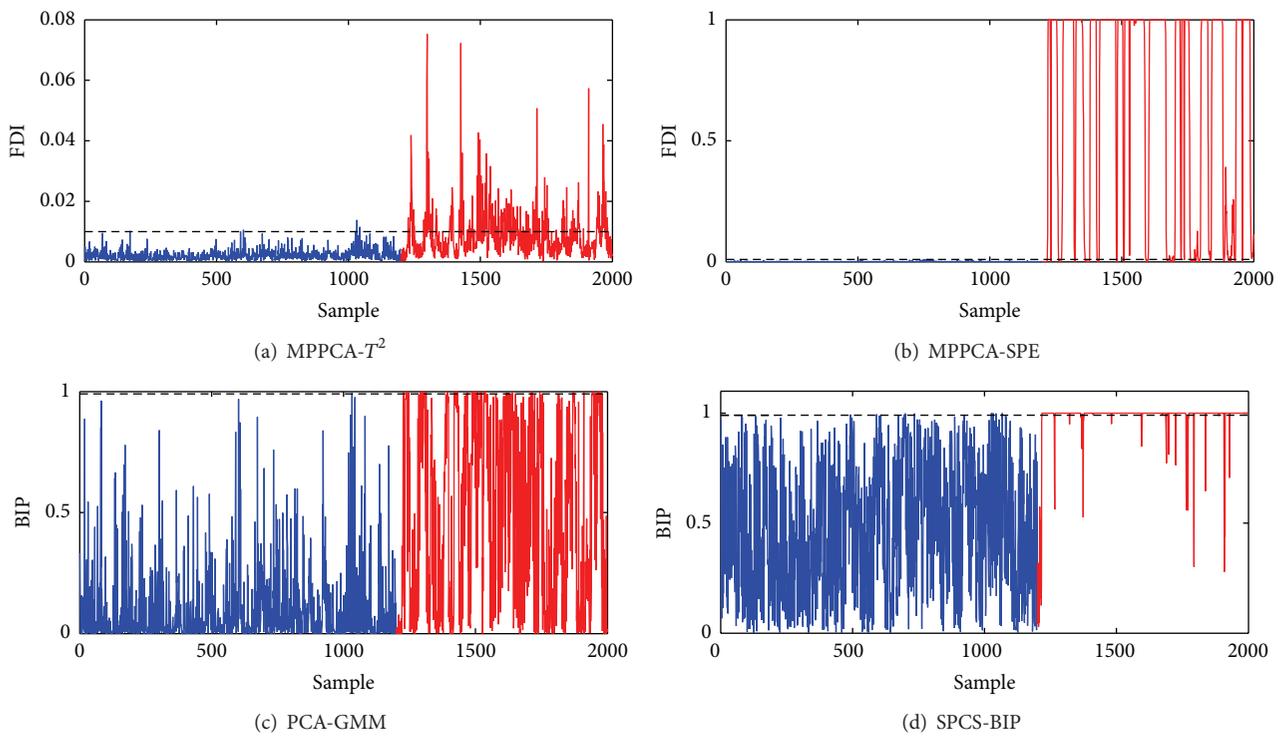


FIGURE 5: Monitoring performances of fault 10 in TEP.

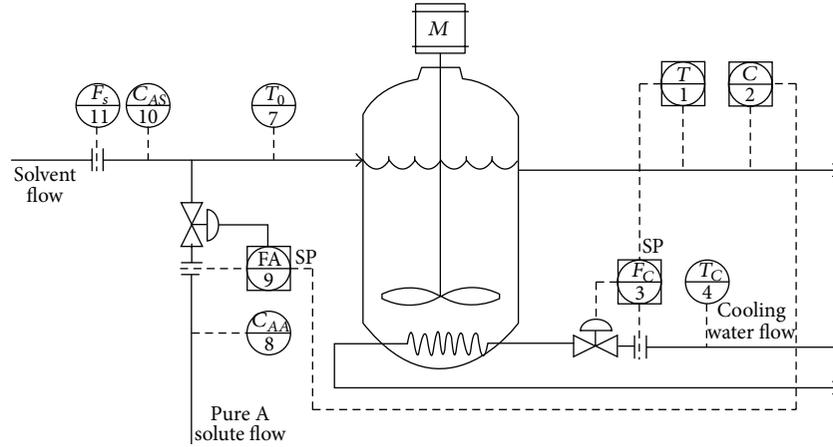


FIGURE 6: Diagram of the CSTR process.

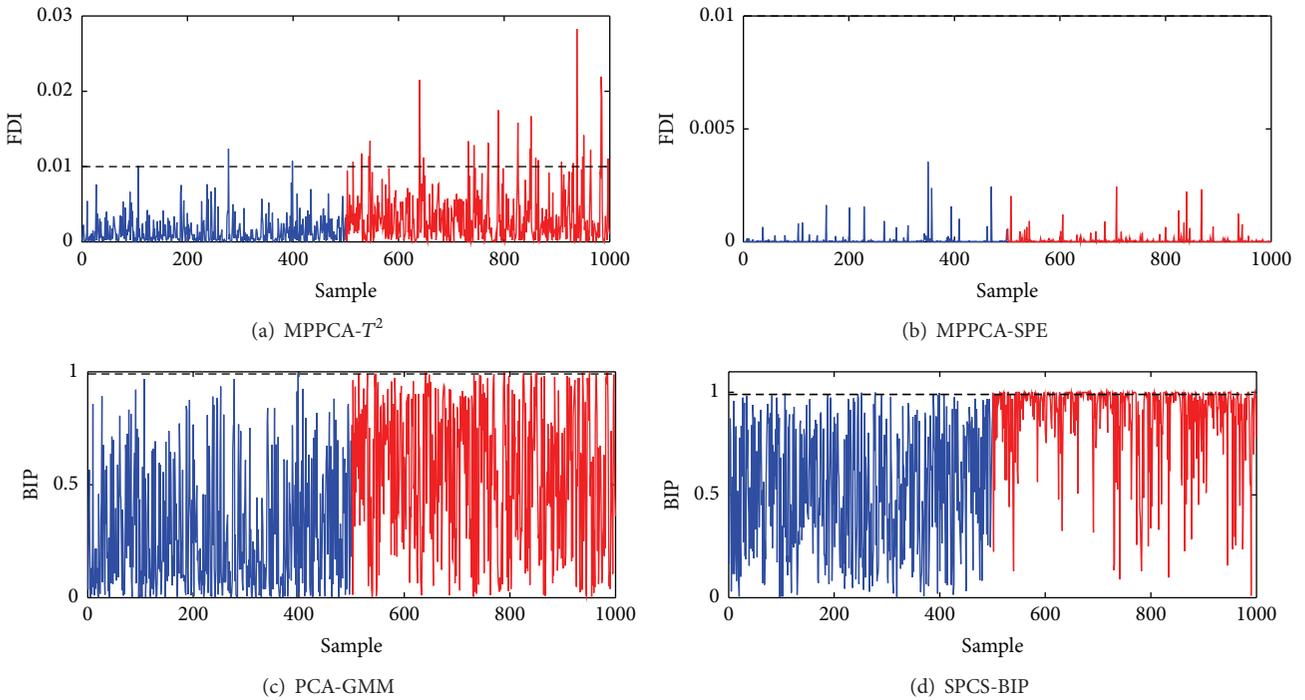


FIGURE 7: Monitoring performance of fault 1 in CSTR.

improvements. Both PCA-GMM and the proposed SPCS are all using BIP. In Figures 8(c) and 8(d), we could hardly see which algorithm is better. However, in Table 4, we could obviously find that the SPCS-BIP is better. Even, compared to MPPCA- T^2 , the proposed algorithm has a little advantage than MPPCA- T^2 .

5. Conclusions

An algorithm using sparse principal component selection and Bayesian inference-based probability (SPCS-BIP) was proposed in this study. Given that the modern industrial processes typically have multiple operating modes, BIP is utilized to compute the posterior probabilities of each monitored sample belonging to the multiple components

and derive an integrated global probabilistic index for fault detection of multimode processes. In each submode, we use the sparse principal component selection to select the key PCs that have the best relation with fault. This algorithm constructs an elastic net regression between all PCs and each sample and then selects PCs according to the nonzero regression coefficients which indicate the discriminative expression of the sample. Finally, the TE and CSTR processes were employed to verify the superiority of the SPCS-BIP algorithm. The monitoring performances of MPPCA, PCA-GMM, and SPCS-BIP methods are discussed compared to those of the MPPCA and PCA-GMM algorithms, and the monitoring performances of the SPCS-BIP algorithm were found to be the best ones among the three algorithms.

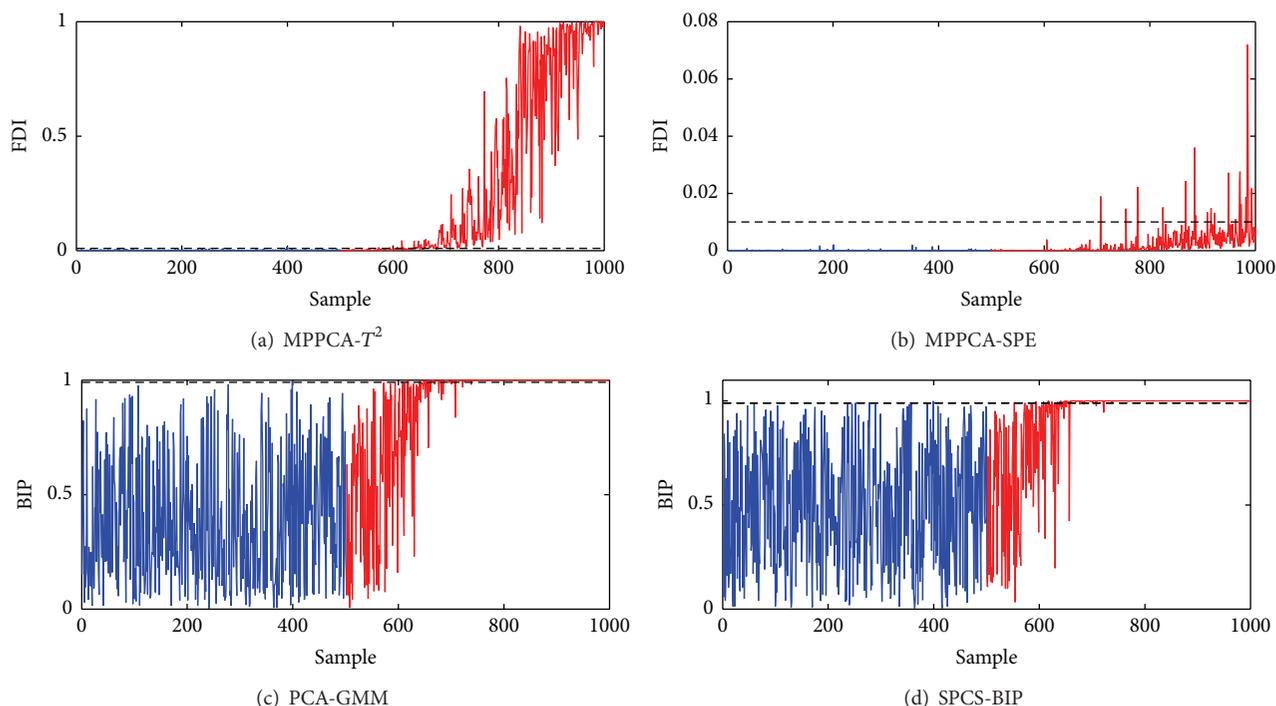


FIGURE 8: Monitoring performance of fault 2 in CSTR.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 61375007) and Shanghai Science and Research Projects (Grant nos. 15JC1400600, 15JC1401700).

References

- [1] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 59, no. 1, pp. 223–234, 2004.
- [2] Z. Ge and Z. Song, "Mixture Bayesian regularization method of PPCA for multimode process monitoring," *AIChE Journal*, vol. 56, no. 11, pp. 2838–2849, 2010.
- [3] D. Kim and I.-B. Lee, "Process monitoring based on probabilistic PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 67, no. 2, pp. 109–123, 2003.
- [4] H. D. Jin, Y.-H. Lee, G. Lee, and C. Han, "Robust recursive principal component analysis modeling for adaptive monitoring," *Industrial and Engineering Chemistry Research*, vol. 45, no. 2, pp. 696–703, 2006.
- [5] C. Zhao and F. Gao, "Fault-relevant Principal Component Analysis (FPCA) method for multivariate statistical modeling and process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 133, pp. 1–16, 2014.
- [6] C. Tong, A. Palazoglu, and X. Yan, "An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding," *Journal of Process Control*, vol. 23, no. 10, pp. 1497–1507, 2013.
- [7] J. Liu and D.-S. Chen, "Operational performance assessment and fault isolation for multimode processes," *Industrial and Engineering Chemistry Research*, vol. 49, no. 8, pp. 3700–3714, 2010.
- [8] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial and Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.
- [9] B. R. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, vol. 44, no. 7, pp. 1596–1610, 1998.
- [10] X. Wang, U. Kruger, and B. Lennox, "Recursive partial least squares algorithms for monitoring complex industrial processes," *Control Engineering Practice*, vol. 11, no. 6, pp. 613–632, 2003.
- [11] Z. Ge and Z. Song, "Process monitoring based on independent Component Analysis-Principal Component Analysis (ICA-PCA) and similarity factors," *Industrial and Engineering Chemistry Research*, vol. 46, no. 7, pp. 2054–2063, 2007.
- [12] Y. Hu, H. Ma, and H. Shi, "Robust online monitoring based on spherical-kernel partial least squares for nonlinear processes with contaminated modeling data," *Industrial and Engineering Chemistry Research*, vol. 52, no. 26, pp. 9155–9164, 2013.
- [13] Y. Ma and H. Shi, "Multimode process monitoring based on aligned mixture factor analysis," *Industrial & Engineering Chemistry Research*, vol. 53, no. 2, pp. 786–799, 2014.
- [14] J.-M. Lee, I.-B. Lee, and C. Yoo, "Statistical process monitoring with independent component analysis," *Journal of Process Control*, vol. 14, no. 5, pp. 467–485, 2004.

- [15] Q. P. He, S. J. Qin, and J. Wang, "A new fault diagnosis method using fault directions in Fisher discriminant analysis," *AIChE Journal*, vol. 51, no. 2, pp. 555–571, 2005.
- [16] C. Lee, S. W. Choi, and I.-B. Lee, "Variable reconstruction and sensor fault identification using canonical variate analysis," *Journal of Process Control*, vol. 16, no. 7, pp. 747–761, 2006.
- [17] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AIChE Journal*, vol. 52, no. 10, pp. 3501–3514, 2006.
- [18] D.-H. Hwang and C. Han, "Real-time monitoring for a process with multiple operating modes," *Control Engineering Practice*, vol. 7, no. 7, pp. 891–902, 1999.
- [19] Z. Ge and Z. Song, "Multimode process monitoring based on Bayesian method," *Journal of Chemometrics*, vol. 23, no. 12, pp. 636–650, 2009.
- [20] Q. P. He, Q. P. He, and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 4, pp. 345–354, 2007.
- [21] S. Natarajan and R. Srinivasan, "Multi-model based process condition monitoring of offshore oil and gas production process," *Chemical Engineering Research and Design*, vol. 88, no. 5-6, pp. 572–591, 2010.
- [22] Y. S. Ng and R. Srinivasan, "An adjoined multi-model approach for monitoring batch and transient operations," *Computers and Chemical Engineering*, vol. 33, no. 4, pp. 887–902, 2009.
- [23] J. Liu and D.-S. Chen, "Fault detection and identification using modified bayesian classification on PCA subspace," *Industrial and Engineering Chemistry Research*, vol. 48, no. 6, pp. 3059–3077, 2009.
- [24] S. J. Zhao, J. Zhang, and Y. M. Xu, "Monitoring of processes with multiple operating modes through multiple principle component analysis models," *Industrial and Engineering Chemistry Research*, vol. 43, no. 22, pp. 7025–7035, 2004.
- [25] J. Yu and S. J. Qin, "Multimode process monitoring with bayesian inference-based finite gaussian mixture models," *AIChE Journal*, vol. 54, no. 7, pp. 1811–1829, 2008.
- [26] Z. Ge and Z. Song, "Maximum-likelihood mixture factor analysis model and its application for process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 102, no. 1, pp. 53–61, 2010.
- [27] Z. Ge, F. Gao, and Z. Song, "Two-dimensional Bayesian monitoring method for nonlinear multimode processes," *Chemical Engineering Science*, vol. 66, no. 21, pp. 5173–5183, 2011.
- [28] Q. Jiang and X. Yan, "Chemical processes monitoring based on weighted principal component analysis and its application," *Chemometrics and Intelligent Laboratory Systems*, vol. 119, pp. 11–20, 2012.
- [29] N. Lu, F. Gao, and F. Wang, "Sub-PCA modeling and on-line monitoring strategy for batch processes," *AIChE Journal*, vol. 50, no. 1, pp. 255–259, 2004.
- [30] G. Diana and C. Tommasi, "Cross-validation methods in principal component analysis: a comparison," *Statistical Methods and Applications*, vol. 11, no. 1, pp. 71–82, 2002.
- [31] I. T. Jolliffe, "A note on the use of principal components in regression," *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 31, no. 3, pp. 300–303, 1982.
- [32] T. Togkalidou, R. D. Braatz, B. K. Johnson, O. Davidson, and A. Andrews, "Experimental design and inferential modeling in pharmaceutical crystallization," *AIChE Journal*, vol. 47, no. 1, pp. 160–168, 2001.
- [33] H. C. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [34] Q. Jiang, X. Yan, and W. Zhao, "Fault detection and diagnosis in chemical processes using sensitive principal component analysis," *Industrial and Engineering Chemistry Research*, vol. 52, no. 4, pp. 1635–1644, 2013.
- [35] A. Arbel, I. H. Rinard, and R. Shinnar, "Dynamics and control of fluidized catalytic crackers. 3. Designing the control system: choice of manipulated and measured variables for partial control," *Industrial and Engineering Chemistry Research*, vol. 35, no. 7, pp. 2215–2233, 1996.
- [36] H. Q. Wang, Z. H. Song, and P. Li, "Fault detection behavior and performance analysis of principal component analysis based process monitoring methods," *Industrial and Engineering Chemistry Research*, vol. 41, no. 10, pp. 2455–2464, 2002.
- [37] V. H. Nguyen and J.-C. Golinval, "Fault detection based on kernel principal component analysis," *Engineering Structures*, vol. 32, no. 11, pp. 3683–3691, 2010.
- [38] S. W. Choi and I.-B. Lee, "Nonlinear dynamic process monitoring based on dynamic kernel PCA," *Chemical Engineering Science*, vol. 59, no. 24, pp. 5897–5908, 2004.
- [39] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, NY, USA, 2001.
- [40] P. Paalanen, J.-K. Kamarainen, J. Ilonen, and H. Kälviäinen, "Feature representation and discrimination based on Gaussian mixture model probability densities-practices and algorithms," *Pattern Recognition*, vol. 39, no. 7, pp. 1346–1358, 2006.
- [41] L. K. Ren and W. M. Lv, "Fault detection via sparse representation for semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 27, no. 2, pp. 252–259, 2014.
- [42] L. Csato and M. Opper, "Sparse representation for gaussian process models," *Advances in Neural Information Processing Systems*, vol. 49, no. 1, pp. 444–450, 2001.
- [43] Z. H. Lai, "Sparse local discriminant projections for discriminant knowledge extraction and classification," *IET Computer Vision*, vol. 6, no. 6, pp. 551–559, 2012.
- [44] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.
- [45] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.
- [46] J. J. Downs and E. F. Vogel, "Plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.
- [47] P. R. Lyman and C. Georgakis, "Plant-wide control of the tennessee Eastman problem," *Computers and Chemical Engineering*, vol. 19, no. 3, pp. 321–331, 1995.
- [48] N. L. Ricker, "Decentralized control of the tennessee eastman challenge process," *Journal of Process Control*, vol. 6, no. 4, pp. 205–221, 1996.
- [49] S. Yoon and J. F. MacGregor, "Fault diagnosis with multivariate statistical models part I: using steady state fault signatures," *Journal of Process Control*, vol. 11, no. 4, pp. 387–400, 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

