



Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Article

Local component based principal component analysis model for multimode process monitoring

Yuan Li^{*}, Dongsheng Yang

Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China



ARTICLE INFO

Article history:

Received 30 June 2020

Received in revised form 18 September 2020

Accepted 24 October 2020

Available online 27 November 2020

Keywords:

Principal component analysis

Finite Gaussian mixture model

Process monitoring

Tennessee Eastman (TE) process

ABSTRACT

For plant-wide processes with multiple operating conditions, the multimode feature imposes some challenges to conventional monitoring techniques. Hence, to solve this problem, this paper provides a novel local component based principal component analysis (LCPCA) approach for monitoring the status of a multimode process. In LCPCA, the process prior knowledge of mode division is not required and it purely based on the process data. Firstly, LCPCA divides the processes data into multiple local components using finite Gaussian mixture model mixture (FGMM). Then, calculating the posterior probability is applied to determine each sample belonging to which local component. After that, the local component information (such as mean and standard deviation) is used to standardize each sample of local component. Finally, the standardized samples of each local component are combined to train PCA monitoring model. Based on the PCA monitoring model, two monitoring statistics T^2 and SPE are used for monitoring multimode processes. Through a numerical example and the Tennessee Eastman (TE) process, the monitoring result demonstrates that LCPCA outperformed conventional PCA and LNS-PCA in the fault detection rate.

© 2020 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights reserved.

1. Introduction

In the industrial process, process monitoring technique is a tool that can be used to enhance the overall efficiency of industrial equipment [1,2]. In recent years, multivariate statistical process monitoring (MSPM) methods have gained more consideration since the increasing availability of process data [3,4]. The most popular method applied to MSPM is the principal component analysis (PCA). Nevertheless, PCA is inappropriate for monitoring processes with complex features. Thus, the PCA extension methods have been developed, including dynamic PCA [5,6], kernel PCA [7,8], multiway PCA [9,10], and other methods [11] to solve the problem of dynamic, nonlinearity, and multi-batch in the processes. For processes with multiple operating conditions, multi-block PCA [12,13], distributed PCA [14,15] and local PCA [16,17] methods have been used for monitoring the status of a multimode process.

The processes with multiple operating conditions have drawn increasing interest in modern industries. Unfortunately, the global PCA is inappropriate for monitoring the multimode processes since it ignores the local process information and the monitoring results are also difficult to interpret [18]. Hence, in multimode processes,

the multi-block approach is proposed to reduce process monitoring complexity and provide accurate monitoring results. MacGregor *et al.* proposed a multi-block partial least square (PLS) method and established monitoring charts for the individual process subsections as well as the entire process [19]. Qin *et al.* analyzed several multi-block PCA and PLS methods in detail [20]. Zhao *et al.* proposed a multiple PCA monitoring method, which used the principal angle to measure the similarities between any PCA models and then determined squared prediction error (SPE) to facilitate process monitoring [21]. Jiang *et al.* developed a multimode process monitoring method based on mutual information-based multi-block PCA, joint probability, and Bayesian inference [22]. However, the multi-block method ignores the process information during the transition period of multiple operating modes. Besides, the knowledge of manually segmented process is always unavailable practically. To solve these issues, mixture models are applied to monitor the status of multimode processes [23]. Yu *et al.* developed a finite Gaussian mixture model (FGMM) to automatically identify the number of Gaussian components. Then the Bayesian inference is applied to derive an integrated global probabilistic index for fault detection [24]. Yu *et al.* proposed a principal components-based Gaussian mixture model (PC-GMM), which is used multiple Gaussian components for handling complex processes with nonlinearity or multimode features [25]. Zhu *et al.* proposed a distributed GMM method to divide sub-block and

^{*} Corresponding author.

E-mail address: li-yuan@mail.tsinghua.edu.cn (Y. Li).

operating mode recognition [26]. Jiang *et al.* proposed a variational Bayesian Gaussian mixture model with canonical correlation analysis strategy [27]. In practice, the above multimode process monitoring methods have been proved to be effective. However, the establishment of these models depends on the reliable mode recognition performance of the mixture model.

Hence, local modeling methods have been developed to monitor the status of a multimode process. He *et al.* proposed a K nearest neighbor-based local modeling method, which uses local distance for monitoring processes [28,29]. However, this method only considers the location information of data and ignores the data direction variation. Ma *et al.* applied the local neighborhood standardization (LNS) strategy to eliminate the influence of multimode features on process control, and then built a PCA model (LNS-PCA) to monitor the entire process [30]. Deng *et al.* proposed a local neighborhood similarity analysis method for monitoring processes [31]. Unfortunately, these methods require numerous computations to search for local information. Meanwhile, the prior knowledge to select the number of neighbors is difficult to obtain.

Basically, a good local model requires local data to be as simple as possible, while the entire process can be described by all neighbor information. Based on these analysis, this paper provides a novel local component based PCA monitoring model (LCPCA). The main contributions of this paper are as follows. First, this paper uses local component standardization to describe the whole multimode processes, which eliminates the impact of multimode properties on the conventional PCA method. Second, by FGMM, LCPCA can automatically identify the number of local components and be used to standardize the historical process data. Third, LCPCA can effectively avoid the problem of local dataset including different mode data and provide suitable data foundation for monitoring process.

The rest of this study is structured as follows. Firstly, PCA and LNS-PCA method are provided in Section 2. Then, the LCS-PCA monitoring method is proposed in Section 3. In Section 4 and Section 5, two case studies are presented through a numerical example and a benchmark of the Tennessee Eastman process. Ultimately, conclusions are made in Section 6.

2. Traditional Method: PCA and LNS-PCA

Suppose that $\mathbf{X} \in \mathbb{R}^{d \times N}$ is a data matrix, where N is the number of samples and d is the number of variables. At first, the covariance of $\mathbf{X}$ can be calculated as follows:

$$\mathbf{C} = \frac{1}{N-1} \mathbf{X} \mathbf{X}^T \quad (1)$$

where $\mathbf{C}$ represents the covariance matrix of $\mathbf{X}$. Then, PCA obtains the loading matrix $\mathbf{P}$ through eigenvalue decomposition of $\mathbf{C}$. After that, PCA decomposes $\mathbf{X}$ as the score $\mathbf{T} \in \mathbb{R}^{K \times N}$, loading $\mathbf{P} \in \mathbb{R}^{d \times K}$, and residual $\mathbf{E} \in \mathbb{R}^{d \times N}$, as follows:

$$\mathbf{X} = \mathbf{P} \mathbf{T} + \mathbf{E} = \mathbf{P} \mathbf{P}^T \mathbf{X} + \mathbf{E} \quad (2)$$

where $\kappa \leq d$ is the number of principal components (PCs). In general, the number of κ can be calculated through cumulative percentage of variance (CPV) method. The columns of $\mathbf{P}$ are actually eigenvectors, which determined through the first κ eigenvalues.

For online process monitoring, T^2 and SPE statistics are used to monitor the principal component subspace (PCS) and the residual subspace (RS), individually. Consider a new sample $\mathbf{x}_{\text{new}} \in \mathbb{R}^d$, the T^2 statistic can be calculated as:

$$T^2 = \mathbf{x}_{\text{new}}^T \mathbf{P} \mathbf{A}^{-1} \mathbf{P}^T \mathbf{x}_{\text{new}} \quad (3)$$

where $\mathbf{A} \in \mathbb{R}^{K \times K}$ is a diagonal matrix indicating the variances of the retained principal component scores. The SPE statistic can be calculated as:

$$SPE = \|\mathbf{x}_{\text{new}} - \hat{\mathbf{x}}_{\text{new}}\|^2 = \|(\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{x}_{\text{new}}\|^2 \quad (4)$$

where $\hat{\mathbf{x}}_{\text{new}}$ denotes the projected sample vector, and $\mathbf{I} \in \mathbb{R}^{d \times d}$ shows an identity matrix. The control limits of T^2 and SPE statistics are determined using the kernel density estimation (KDE) method [11,32].

To solve the problem caused by the multimode features, Ma *et al.* developed the LNS-PCA monitoring method, which utilized neighborhood mean and standard deviation to standardize the samples [30,33]. The detail of LNS-PCA is as follows.

For a new sample $\mathbf{x}_i$, the π nearest neighbors of $\mathbf{x}_i$ in training data $\mathbf{X}$ can be determined as:

$$N(\mathbf{x}_i) = [\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, \dots, \mathbf{x}_i^{(\pi)}]^T \in \mathbb{R}^{\pi \times d} \quad (5)$$

where $\mathbf{x}_i^{(\pi)}$ represents the π -th nearest neighbor of $\mathbf{x}_i$.

As the local neighborhood dataset become available, the LNS method is performed as follows:

$$\mathbf{z}_i = \frac{\mathbf{x}_i - \text{me}(N(\mathbf{x}_i))}{\text{st}(N(\mathbf{x}_i))} \quad (6)$$

where $\text{me}(N(\mathbf{x}_i))$ and $\text{st}(N(\mathbf{x}_i))$ denote the mean and standard deviation of neighborhood dataset.

In comparison with PCA, using LNS strategy makes PCA monitoring the multimode process effectively. However, the number of neighbors affects the monitoring performance of LNS-PCA in the multimode processes. Meanwhile, LNS-PCA utilizes the Euclidean distance to find the local neighborhood, which may cause the neighborhood dataset to still present the multimode features.

3. LCPCA for Process Monitoring

In this section, we will introduce a novel LCPCA method, which is a powerful monitoring technique. Firstly, LCPCA does not required the process prior knowledge of mode division and it is based on the process data, purely. Secondly, the proposed method can avoid the problem which the local neighborhood data includes the multiple mode. Thirdly, LCPCA provides a suitable data foundation for calculating monitoring statistics. The main idea of LCPCA is represented as follows.

Suppose $\mathbf{X} \in \mathbb{R}^{d \times N}$ is a data matrix with N samples and d variables from multimode processes. Firstly, using the FGMM, the multimode process data is divided into the local components in LCPCA. The probability density of $\mathbf{X}$ can be determined as:

$$p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^N \sum_{k=1}^K \omega_k \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (7)$$

where K denotes the number of local components in FGMM. ω_k represents the weight of each local component, and it satisfies $\sum_{k=1}^K \omega_k = 1$. $\mathcal{N}(\cdot)$ indicates the Gaussian probability function. $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are a mean vector and a covariance matrix of the k -th local component, respectively. The unknown parameters $\boldsymbol{\theta} = \{\{\omega_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\}, \dots, \{\omega_K, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K\}\}$ should be estimated to establish the FGMM model. The procedure of parameter estimation is as follows.

The first step is:

$$p^{(s)}(\mathbf{C}_k^{(s)}|\mathbf{x}_i) = \frac{\omega_k^{(s)} \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k^{(s)}, \boldsymbol{\Sigma}_k^{(s)})}{\sum_{j=1}^K \omega_j^{(s)} \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_j^{(s)}, \boldsymbol{\Sigma}_j^{(s)})} \quad (8)$$

where $p^{(s)}(C_k^{(s)}|\mathbf{x}_i)$ is the posterior probability of i -th sample within the k -th local component $C_k^{(s)}$ at the s -th iteration.

The second step is:

$$\mu_k^{(s+1)} = \frac{\sum_{i=1}^N p^{(s)}(C_k^{(s)}|\mathbf{x}_i) \mathbf{x}_i}{\sum_{i=1}^N p^{(s)}(C_k^{(s)}|\mathbf{x}_i)} \quad (9)$$

$$\Sigma_k^{(s+1)} = \frac{\sum_{i=1}^N p^{(s)}(C_k^{(s)}|\mathbf{x}_i) (\mathbf{x}_i - \mu_k^{(s+1)}) (\mathbf{x}_i - \mu_k^{(s+1)})^T}{\sum_{i=1}^N p^{(s)}(C_k^{(s)}|\mathbf{x}_i)} \quad (10)$$

$$\omega_k^{(s+1)} = \frac{\max\{0, (\sum_{i=1}^N p^{(s)}(C_k^{(s)}|\mathbf{x}_i)) - \frac{V}{2}\}}{\sum_{j=1}^K \max\{0, (\sum_{i=1}^N p^{(s)}(C_k^{(s)}|\mathbf{x}_i)) - \frac{V}{2}\}} \quad (11)$$

where $\mu_k^{(s+1)}$, $\Sigma_k^{(s+1)}$, and $\omega_k^{(s+1)}$ are a mean vector, a covariance matrix, and a weight of component $C_k^{(s)}$, respectively. $V = \frac{1}{2}d^2 + \frac{3}{2}d$ denotes the number of scalar parameters specifying each Gaussian component. After estimating the unknown parameters, FGMM can be successfully constructed. Herein, we use a simple multimode case to clearly describe the procedure steps of proposed method [34]. As shown in Fig. 1, the process data is divided into 6 local components by FGMM.

Next, LCPCA computes the posterior probability for each sample $\mathbf{x}$ belonging to each Gaussian component as follows:

$$p(C_k|\mathbf{x}) = \frac{\omega_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)}{\sum_{j=1}^K \omega_j \mathcal{N}(\mathbf{x}|\mu_j, \Sigma_j)} \quad (12)$$

According to the posterior probability, the process data $\mathbf{X}$ can be divided into multiple local component $n(\mathbf{x}_k)$ as follows:

$$\mathbf{X} = \{n(\mathbf{x}_1), n(\mathbf{x}_2), \dots, n(\mathbf{x}_k)\} \quad (13)$$

where $n(\mathbf{x}_k)$ represents the k -th local component. In this step, the samples of local component overlap region can be divided by calculating the posterior probability. We take one of the mode as an example. As shown in Fig. 2, each sample in local component A and B is distinguished by different symbols. It can be clearly seen that each sample can be described using to the corresponding local component.

Then, LCPCA standardizes each sample of local component $n(\mathbf{x}_k)$ as follows:

$$\hat{n}(\mathbf{x}_k) = \frac{\{\mathbf{x}_i^{(k)} - m[n(\mathbf{x}_k)]\}}{s[n(\mathbf{x}_k)]} \quad (14)$$

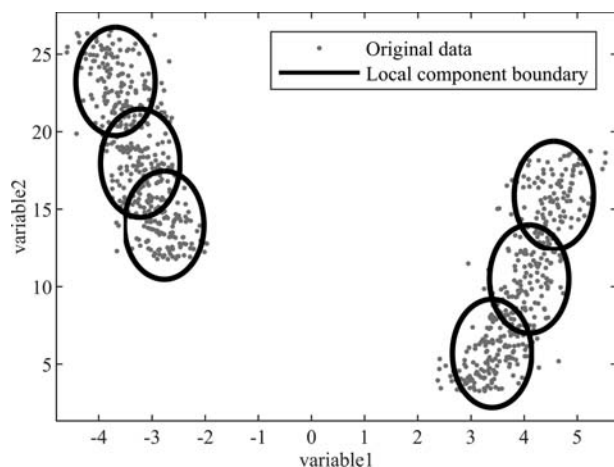


Fig. 1. Scatter plots of the local component.

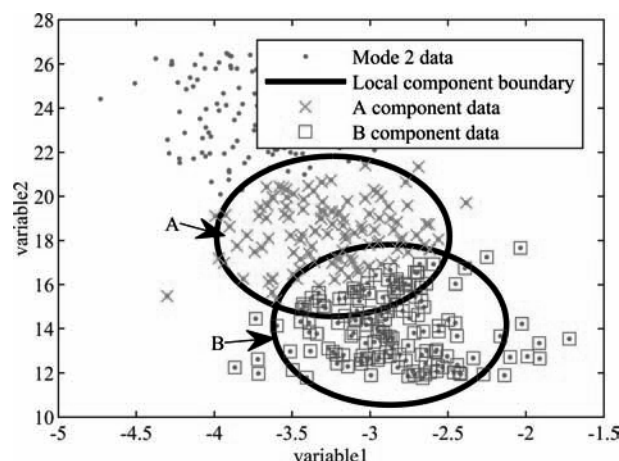


Fig. 2. Scatter plots of the sample contained in each local component.

where $\hat{n}(\mathbf{x}_k)$ denotes the local component, which has been standardized. $\mathbf{x}_i^{(k)}$ represents the i -th sample in the k -th local component. $m[n(\mathbf{x}_k)]$ and $s[n(\mathbf{x}_k)]$ are a mean vector and a standard deviation vector of the k -th local component. Thus, the process data $\mathbf{X}$ can be reconstructed as:

$$\hat{\mathbf{X}} = \{\hat{n}(\mathbf{x}_1), \hat{n}(\mathbf{x}_2), \dots, \hat{n}(\mathbf{x}_k)\} \quad (15)$$

Take two components have different distribution as example, as shown in Fig. 3. Through component information (mean and standard deviation), LCPCA make the standardized component samples $\hat{n}(\mathbf{x}_k)$ close to the origin.

Finally, the PCA model is built to monitor the standardized process data $\hat{\mathbf{X}}$. According to Fig. 4, the process data $\hat{\mathbf{X}}$ have been already normalized to a single distribution and eliminated the multimode features through Eq. (14). Furthermore, LCPCA makes the PCA monitoring model robust performance increasing.

For a new observation sample $\mathbf{x}_*$, LCPCA is first to compute its posterior probability, and determines it belongs to which local component $n(\mathbf{x}_k)$. Then, the local information of $n(\mathbf{x}_k)$ is used to standardize $\mathbf{x}_*$ through Eq. (14). Finally, LCPCA used PCA model to monitor the change in principal component subspace (PCS) and residual subspace (RS) of $\hat{\mathbf{x}}_*$. LCPCA includes two parts, the details of offline modeling and online monitoring are as follows.

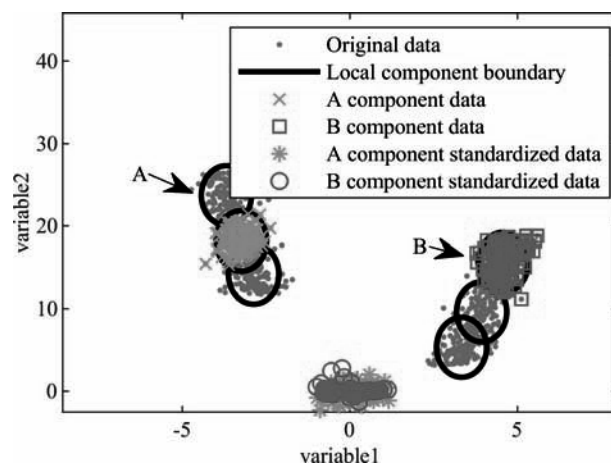


Fig. 3. Scatter plots of the standardized samples of A and B local component.

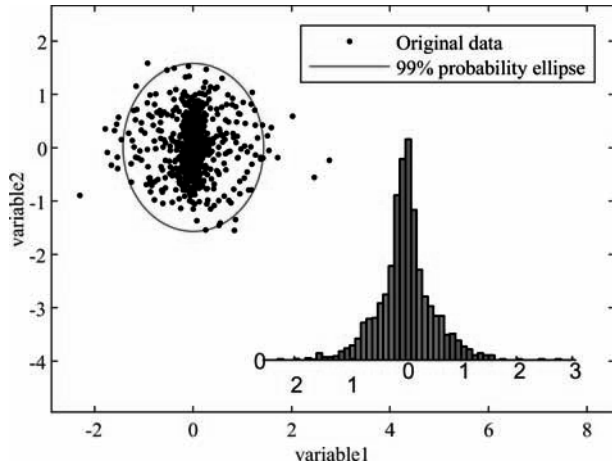


Fig. 4. Scatter plots of standardized samples and its variable distribution histogram in LCPCA.

Part 1: Offline modeling

- (1) Collect the process data $\mathbf{X}$ under normal multimode operation conditions;
- (2) Decompose $\mathbf{X}$ into multiple local component $n(\mathbf{x}_k)$ using FGMM technique;
- (3) Use the local information (such as $m[n(\mathbf{x}_k)]$ and $s[n(\mathbf{x}_k)]$) to standardize each sample in $n(\mathbf{x}_k)$;
- (4) Acquire standardized process data $\hat{\mathbf{X}}$;
- (5) Build the PCA model on the process data matrix $\hat{\mathbf{X}}$;
- (6) Calculate the T^2 and SPE statistics in PCS and RS, respectively;
- (7) Determine the control limits CL_{T^2} and CL_{SPE} using KDE.

Part 2: Online monitoring

- (1) Acquire a new sample $\mathbf{x}_{\text{new}}$, calculate its posterior probability to determine it belongs to which local component;
- (2) Use the local information to standardize $\mathbf{x}_{\text{new}}$, and get $\hat{\mathbf{x}}_{\text{new}}$;
- (3) Map $\hat{\mathbf{x}}_{\text{new}}$ to the PCS and RS, and then calculate the T^2_{new} and SPE_{new} statistics;

Compare the T^2_{new} and SPE_{new} statistics with the control limits CL_{T^2} and CL_{SPE} , respectively. If $T^2_{\text{new}} > CL_{T^2}$ or $SPE_{\text{new}} > CL_{SPE}$, $\mathbf{x}_{\text{new}}$ is identified as fault sample, otherwise, it is identified as normal sample.

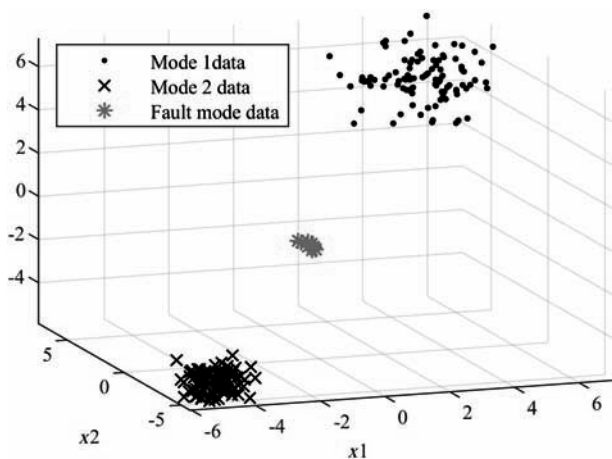


Fig. 5. Scatter plots of the numerical example data.

4. Numerical Case Study

In this section, a simple numerical example is used to the simulation experiment. The simulation data are generated from the following process system:

$$\mathbf{X} = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3] \quad (16)$$

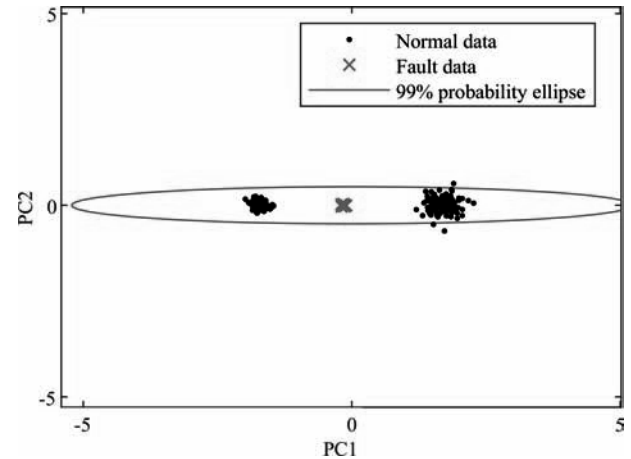


Fig. 6. Scatter plots of the samples in PCS.

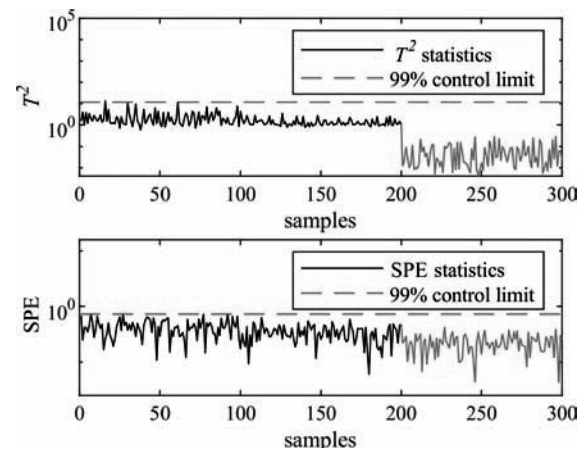


Fig. 7. Monitoring results for the numerical example using PCA.

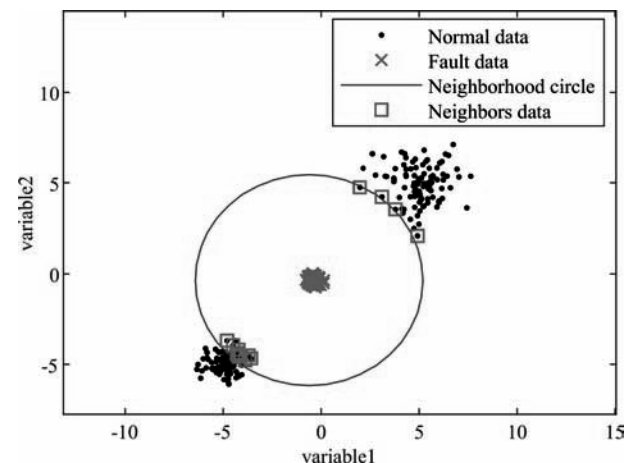


Fig. 8. Scatter plots of the neighborhood of the fault sample in LNS-PCA.

where $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$ represent the signal sources. Through adjusting the signal sources $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ to generate different operating modes as follows:

Mode 1: $\mathbf{s}_1 \sim \mathcal{N}(5, 1)$, $\mathbf{s}_2 \sim \mathcal{N}(5, 1)$, $\mathbf{s}_3 \sim \mathcal{N}(5, 1)$;

Mode 2: $\mathbf{s}_1 \sim \mathcal{N}(-5, 0.5)$, $\mathbf{s}_2 \sim \mathcal{N}(-5, 0.5)$, $\mathbf{s}_3 \sim \mathcal{N}(-5, 0.5)$;

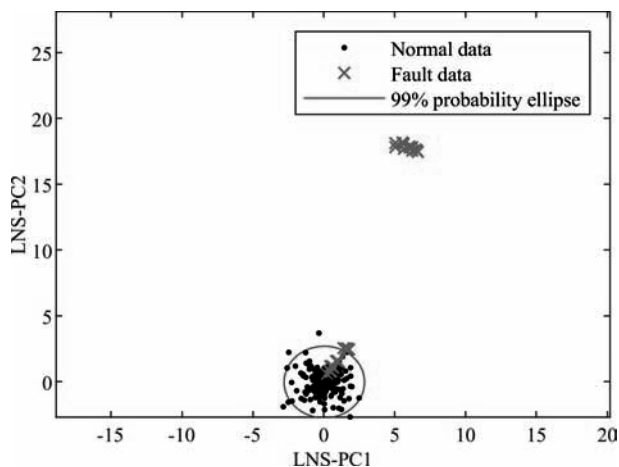


Fig. 9. Scatter plots of the standardized samples in PCS of LNS-PCA.

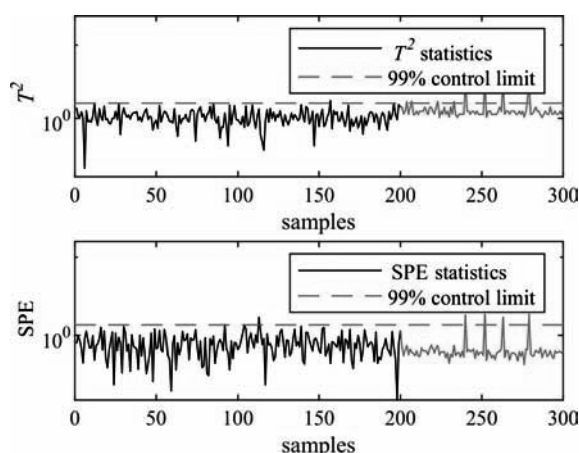


Fig. 10. Monitoring results for the numerical example using LNS-PCA.

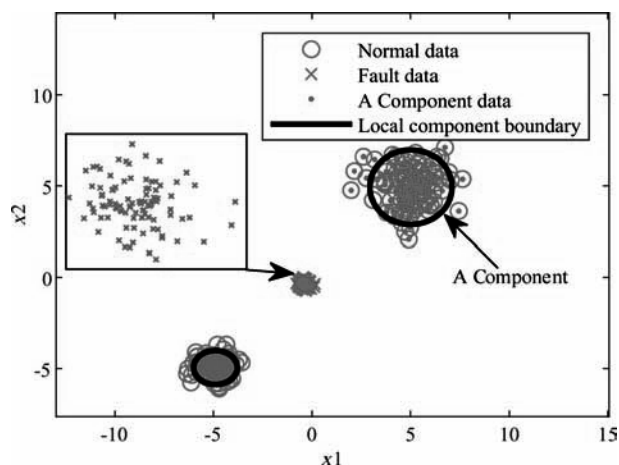


Fig. 11. Scatter plots of the sample of a local component in LCPCA.

In this simulation, we used this process system to generate 100 samples under each mode, and a total of 200 samples are used for off-line modeling. Then, two datasets are generated as the testing data, and the 100 fault samples are generated as follows:

- (1) In mode 1, the system runs at mode 1 to generate 100 samples;
- (2) In mode 2, the system runs at mode 2 to generate 100 samples;
- (3) In fault mode, the signal sources are follow $\mathbf{s}_1 \sim \mathcal{N}(-0.4, 0.15)$, $\mathbf{s}_2 \sim \mathcal{N}(-0.4, 0.15)$, $\mathbf{s}_3 \sim \mathcal{N}(-0.4, 0.15)$ to generate 100 fault samples.

In the end, a total of 300 testing samples containing mode 1, mode 2 and fault mode are used to verify PCA, LNS-PCA and the proposed LCPCA method. All of the testing data are shown in Fig. 5.

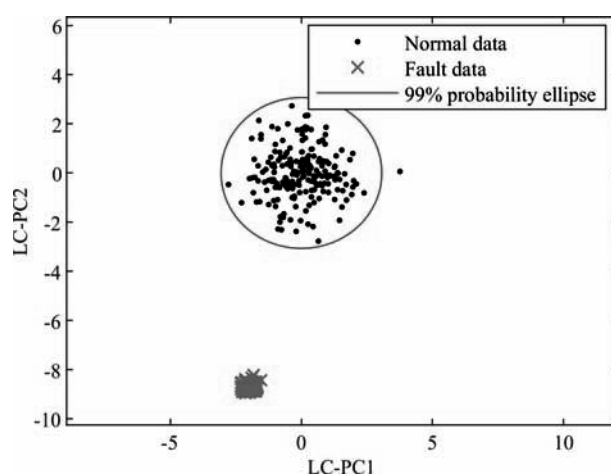


Fig. 12. Scatter plots of the standardized samples in PCS of LCPCA.

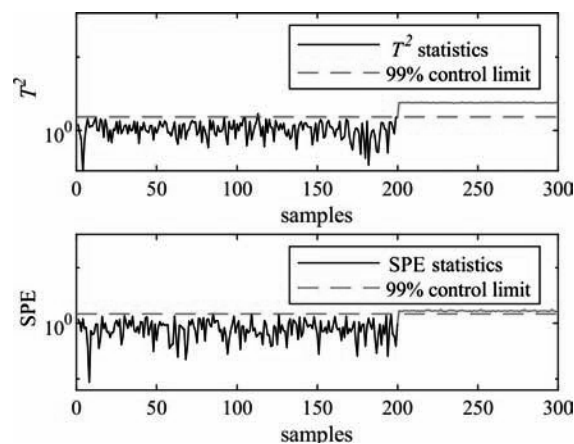


Fig. 13. Monitoring results for the numerical example using LCPCA.

Table 1
Fault detection rates of PCA, LNS-PCA, and LCPCA

Method	Fault detection rates	
	T^2	SPE
PCA	0	0
LNS-PCA	0.09	0.04
LCPCA	1	1

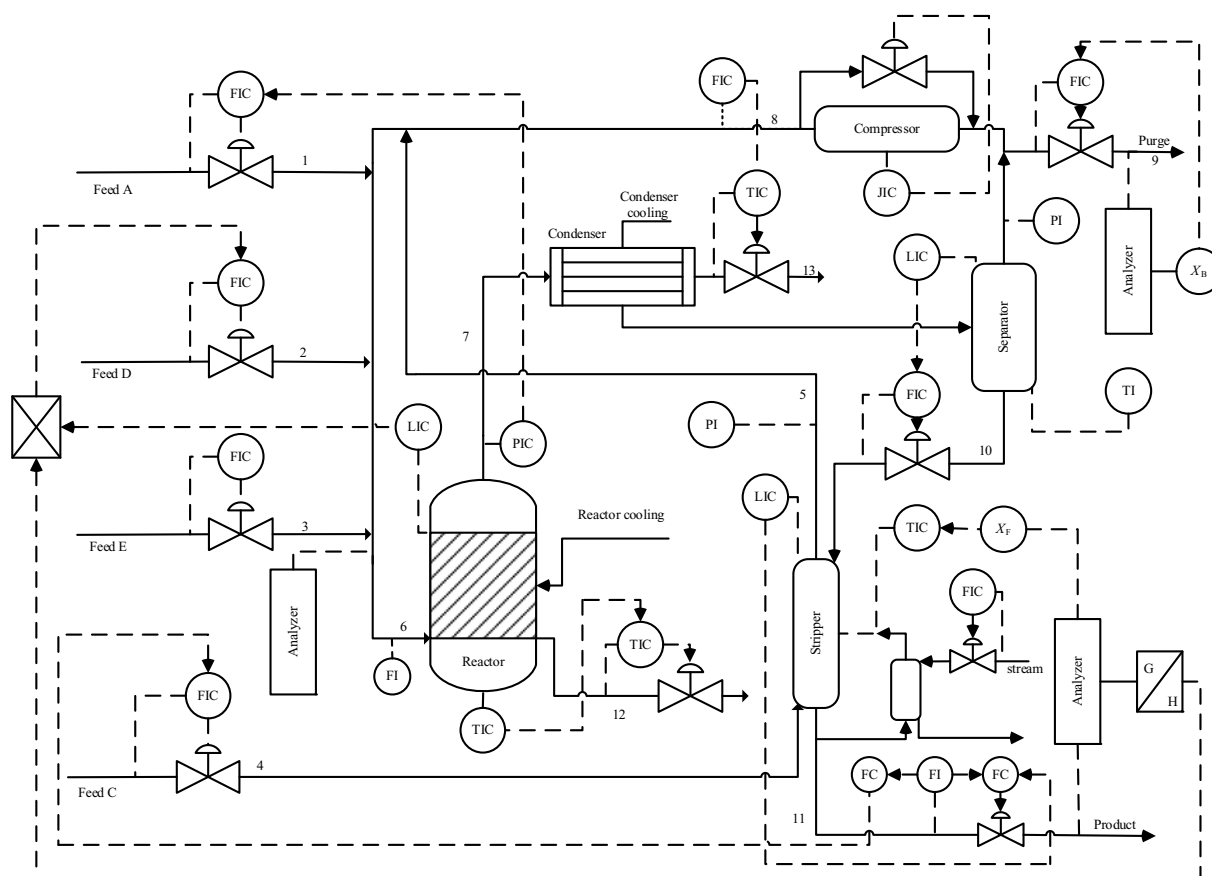


Fig. 14. Tennessee Eastman Benchmark process.

In PCA, the number of PCs is calculated as 2 through the 85% CPV. In LNS-PCA, the number of neighbors is chosen as 15, and the number of PCs is calculated as 2. In LCPCA, the number of local component is automatic determined by FGMM, and the number of PCs is calculated as 2 through 85% CPV. Moreover, we specify the control limits as 99% for each method. In each monitoring result figure, the red-solid lines denote the fault samples and the red-dotted lines reflect the 99% control limits.

According to Fig. 6, PCA cannot extract multimode features in this process system. In comparison with the 99% probability ellipse of normal distribution, it can be clearly seen that all fault samples are still contained in the ellipse. Hence, based on Fig. 7, the monitoring results shows that all fault samples are not detected.

In LNS-PCA, the standardized results of LNS will be affected by the neighborhood of fault sample. If the number of neighbors are small, the neighborhood data may unable to describe the actual process information. In the other hand, if the number of neighbors are huge, the neighborhood data may include the multiple modes, which makes the neighborhood standard deviation inflated as shown in Fig. 8. Meanwhile, according to the Fig. 9, the most stan-

dardized fault samples cannot be separated from the normal samples. This is because the fault samples minus the neighborhood means, then divides by the inflated standard deviation of its neighborhood is around 0. Hence, as shown in Fig. 10, most fault samples are not detected in LNS-PCA.

In LCPCA, according to Fig. 11, a probability model is used to divide the data into multiple local component and find the corre-

Table 3
Process variables used for process monitoring in TE process

No.	Variables	No.	Variables
X1	A feed	X18	Stripper temperature
X2	D feed	X19	Stripper steam flow
X3	E feed	X20	Compressor work
X4	A and C feed	X21	Reactor cooling water outlet temperature
X5	Recycle flow	X22	Separator cooling water outlet temperature
X6	Reactor feed rate	X23	D feed flow valve
X7	Reactor pressure	X24	E feed flow valve
X8	Reactor level	X25	A feed flow valve
X9	Reactor temperature	X26	A and C feed flow valve
X10	Purge rate	X27	Compressor recycle valve
X11	Product separator temperature	X28	Purge valve
X12	Product separator level	X29	Separator pot liquid flow valve
X13	Product separator pressure	X30	Stripper liquid product flow valve
X14	Product separator underflow	X31	Stripper steam valve
X15	Stripper level	X32	Reactor cooling water flow
X16	Stripper pressure	X33	Condenser cooling water flow
X17	Stripper underflow		

Table 2
Six operation modes in TE process

Mode	G/H mass ration	Production rate
1	50/50	7038 kg·h ⁻¹ G and 7038 kg·h ⁻¹ H
2	10/90	1408 kg·h ⁻¹ G and 12669 kg·h ⁻¹ H
3	90/10	10000 kg·h ⁻¹ G and 1111 kg·h ⁻¹ H
4	50/50	Maximum
5	10/90	Maximum
6	90/10	Maximum

Table 4
Induced process faults in TE process

No.	Process variable	Type
IDV1	A/C feed ratio, B composition constant	Step
IDV2	B composition, A/C ration constant	Step
IDV3	D feed temperature	Step
IDV4	Reactor cooling water inlet temperature	Step
IDV5	Condenser cooling water inlet temperature	Step
IDV6	A feed loss	Step
IDV7	C header pressure loss-reduced availability	Step
IDV8	A, B, and C feed composition	Random variation
IDV9	D feed temperature	Random variation
IDV10	C feed temperature	Random variation
IDV11	Reactor cooling water inlet temperature	Random variation
IDV12	Condenser cooling water inlet temperature	Slow drift
IDV13	Reaction kinetics	Sticking
IDV14	Reactor cooling water valve	Sticking
IDV15	Condenser cooling water valve	Sticking
IDV16	Unknown	Unknown
IDV17	Unknown	Unknown
IDV18	Unknown	Unknown
IDV19	Unknown	Unknown
IDV20	Unknown	Unknown
IDV21	Unknown	Unknown

sponding component for each sample. Meanwhile, this finding indicates that all fault samples belong to the same local component. Since each sample can only be described by one local component, which can avoid the problem of the neighborhood data including multiple modes in LNS-PCA. In comparison with LNS-PCA, LCPCA can successfully maintain the deviation degree among the normal and fault samples using the local component standardization, as shown in Fig. 12. Hence, LCPCA can detect most fault samples effectively as presented in Fig. 13. The monitoring results of PCA, LNS-PCA, and LCPCA are listed in Table 1. The numerical case illustrates LCPCA possesses significant fault detection performance in multimode process.

5. Tennessee Eastman Process

As a famous plant-wide process, the Tennessee Eastman (TE) benchmark is widely used for evaluating process monitoring methods. Downs and Vogel proposed a prototype problem based on the Tennessee Eastman chemical company [35]. The TE process com-

prises five basic units: a reactor, a condenser, a vapor–liquid separator, a recycle compressor, and a product stripper, the systematic illustration as shown in Fig. 14. The TE process includes 12 manipulated variables and 41 measurement variables. There is composed of six different operating modes that have different G/H mass ratios. A detailed description of the modes is listed in Table 2. Additional details on the TE process can be found in Ref [35].

In this simulation, 33 process variables are selected for monitoring, as illustrated in Table 3. The operating Mode 1 and Mode 3 are used to simulate multimode processes. A set of different types of 21 faults is simulated in each mode of TE process as listed in Table 4. To establish the monitoring model, 961 normal samples under each mode are collected. Furthermore, 160 normal samples and 801 fault samples under each mode compose the testing data, and a total of 1922 samples are used to evaluate the efficiency of the proposed monitoring method. In PCA, the number of PCs is determined as 1 through the 85% CPV. In LNS-PCA, the number of neighbors is set as 30 according to the Ref [30], and the number of PCs is set as 18 through the 85% CPV. In LCPCA, the number of local component is automatically determined using the FGMM, and the number of PCs is set as 23 through the 85% CPV. The confidence level of each method is set as 99%. Table 5 illustrates the fault detection results of PCA, LNS-PCA, and LCPCA. It can be found that LCPCA outperforms conventional PCA and LNS-PCA technique in the performance of monitoring. The Fault 10 is selected to demonstrate the effectiveness of the proposed method.

Fault 10 is generated by random variation in C feed temperature. Fig. 15(a) represents the probability plots of the first PCs in the PCA method. This finding indicates that the shifts in the operating mode cause significant differences in process information and then make the process data include multiple modes features. Meanwhile, the conventional PCA method based on the single distribution assumption performs poorly in this case. Hence, as shown in Fig. 15(b), PCA cannot detect most of fault samples. According to the Fig. 16(a), through the LNS-PCA method, the process variables basically follow a Gaussian distribution. However, as mentioned above, LNS-PCA cannot avoid the problem caused by the neighborhood data including multiple modes. Thus, a part of fault samples is not detected as shown in Fig. 16(b). Compared with PCA and LNS-PCA, the LCPCA method can use local component information to standardize process data and make the process variables follow a Gaussian distribution, as shown in Fig. 17(a). Therefore, accord-

Table 5
The fault detection rates of PCA, LNS-PCA, and LCPCA

Fault No.	PCA		LNS-PCA		LCPCA	
	T^2	SPE	T^2	SPE	T^2	SPE
IDV1	0.5	0.99	1	1	1	1
IDV2	0.48	0.57	0.99	0.98	0.99	0.99
IDV3	0.01	0.01	0.01	0.04	0.1	0.1
IDV4	0.5	1	1	1	1	1
IDV5	0.04	0.48	0.5	0.52	0.52	0.53
IDV6	0.52	1	1	1	1	1
IDV7	0.37	0.22	1	1	1	1
IDV8	0.41	0.94	0.98	0.97	0.98	0.98
IDV9	0.01	0.04	0.03	0.1	0.24	0.17
IDV10	0.02	0.04	0.69	0.9	0.95	0.86
IDV11	0.19	0.45	0.74	0.92	0.96	0.92
IDV12	0.16	0.51	0.58	0.62	0.79	0.79
IDV13	0.3	0.9	0.95	0.95	0.95	0.95
IDV14	0.38	0.67	0.98	0.99	1	0.99
IDV15	0.02	0.01	0.01	0.06	0.1	0.11
IDV16	0.01	0.01	0.01	0.03	0.09	0.06
IDV17	0.21	0.69	0.92	0.94	0.94	0.9
IDV18	0.2	0.49	0.78	0.85	0.89	0.9
IDV19	0.03	0.5	0.98	1	1	0.99
IDV20	0.29	0.84	0.89	0.95	0.95	0.93
IDV21	0.01	0.01	0.001	0.03	0.1	0.07

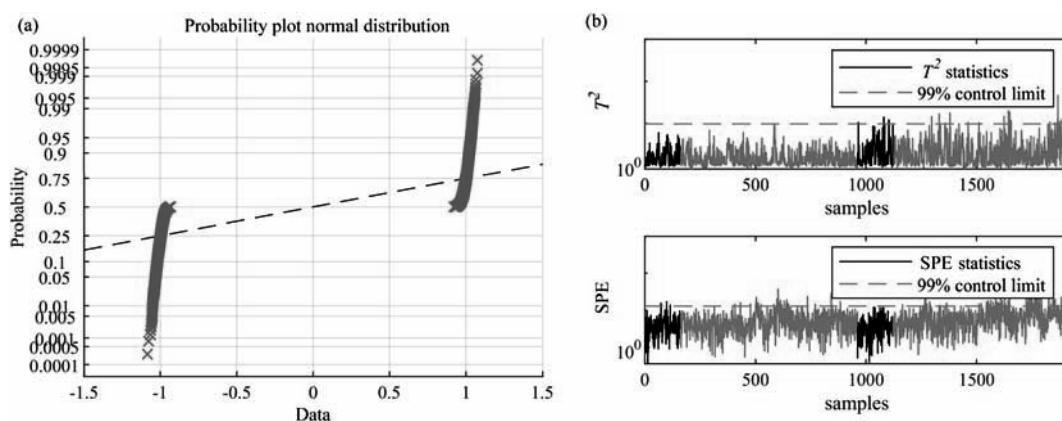


Fig. 15. (a) Probability plots of the first PCs in PCA (b) Monitoring results for TE process using PCA.

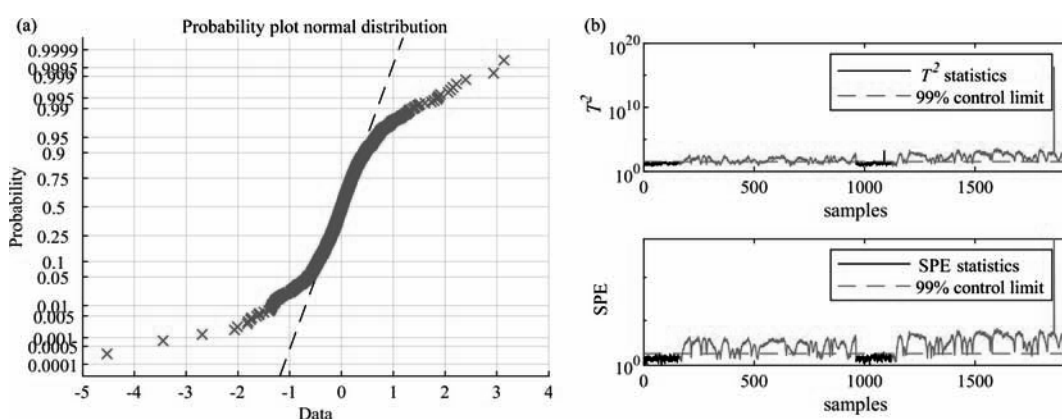


Fig. 16. (a) Probability plots of the first PCs in LNS-PCA (b) Monitoring results for TE process using LNS-PCA.

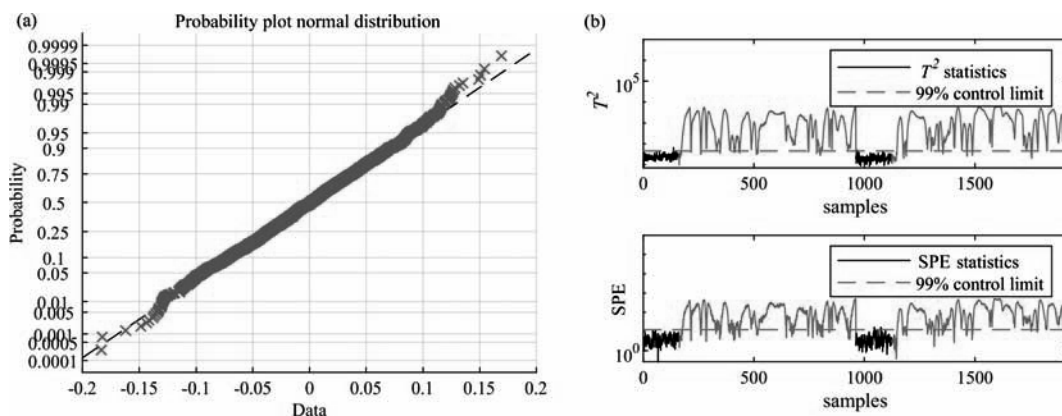


Fig. 17. (a) Probability plots of the first PCs in LCPCA (b) Monitoring results for TE process using LCPCA.

ing to the Fig. 17(b) the monitoring results indicate the LCPCA plays a more effective role in monitoring the multimode process.

6. Conclusions

In this paper, the local component based principal component analysis method is proposed to eliminate the multimode features and then monitor the status of processes. The three features of LCPCA are summarized as follows. First, LCPCA can eliminate the impact of multimode properties in PCA. Second, LCPCA can efficiently divide the process data, and automatically select the number of the local component without prior knowledge. Third,

proposed method can avoid the problem of the local neighborhood data including multiple modes. The validity and superiority of the LCPCA are verified by a multimode numerical case and a benchmark of the Tennessee Eastman process. The application results demonstrate the superior performance of LCPCA compared with PCA and LNS-PCA in the multimode process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge that the National Natural Science Foundation of China (61673279).

References

- [1] X.M. Zhang, M. Kano, M. Tani, J. Mori, J. Ise, K. Harada, Prediction and causal analysis of defects in steel products: Handling nonnegative and highly overdispersed count data, *Control Eng. Pract.* 95 (2020) 104258.
- [2] Y. Zheng, H.J. Fang, H.O. Wang, Takagi-Sugeno fuzzy model-based fault detection for networked control systems with markov delays, *IEEE Trans. Syst. Man and Cybernetics Part B Cybernetics* 36 (3) (2006) 924–929.
- [3] X. Ma, Y.B. Si, Z.Y. Yuan, Y.H. Qin, Y.Q. Wang, Multi-step dynamic slow feature analysis for industrial process monitoring, *IEEE Trans. Inst. Measur.* 69 (12) (2020) 9535–9548.
- [4] Y.B. Si, Y.Q. Wang, D.H. Zhou, Key-performance-indicator-related process monitoring based on improved kernel partial least squares, *IEEE Trans. Ind. Electron.* 68 (3) (2021) 2626–2636.
- [5] Y. Tao, H.B. Shi, B. Song, S. Tan, A novel dynamic weight principal component analysis method and hierarchical monitoring strategy for process fault detection and diagnosis, *IEEE Trans. Ind. Electron.* 67 (9) (2020) 7994–8004.
- [6] K. Wang, J.H. Chen, Z.H. Song, Performance analysis of dynamic PCA for closed-loop process monitoring and its improvement by output oversampling scheme, *IEEE Trans. Control Syst. Technol.* 27 (1) (2019) 378–385.
- [7] Y.W. Zhang, S. Li, Z.Y. Hu, Improved multi-scale kernel principal component analysis and its application for fault detection, *Chem. Eng. Res. Des.* 90 (9) (2012) 1271–1280.
- [8] J.M. Lee, C.K. Yoo, S.W. Choi, P.A. Vanrolleghem, I.B. Lee, Nonlinear process monitoring using kernel principal component analysis, *Chem. Eng. Sci.* 59 (1) (2004) 223–234.
- [9] P. Nomikos, J.F. MacGregor, Monitoring batch processes using multiway principal component analysis, *AIChE J.* 40 (1994) 1361–1375.
- [10] F. Pozo, Y. Vidal, O. Salgado, Wind turbine condition monitoring strategy through multiway PCA and multivariate inference, *Energies* 11 (4) (2018) 749.
- [11] Z.Q. Ge, Z.H. Song, F.R. Gao, Review of recent research on data-based process monitoring, *Ind. Eng. Chem. Res.* 52 (10) (2013) 3543–3562.
- [12] Y.Q. Chang, R.X. Ma, L.P. Zhao, F.L. Wang, S. Wang, Online operating performance evaluation for the plant-wide industrial process based on a three-level and multi-block method, *Can. J. Chem. Eng.* 97 (S1) (2019) 1371–1385.
- [13] A. Jaworski, K. Wikiel, H. Wikiel, Application of multiblock and hierarchical PCA and PLS models for analysis of AC voltammetric data, *Electroanalysis* 17 (15–16) (2005) 1477–1485.
- [14] Z.Q. Ge, Z.H. Song, Distributed PCA model for plant-wide process monitoring, *Ind. Eng. Chem. Res.* 52 (5) (2013) 1947–1957.
- [15] M. Rezamand, M. Kordestani, R. Carrière, D.-S.-K. Ting, M. Saif, A new hybrid fault detection method for wind turbine blades using recursive PCA and wavelet-based PDF, *IEEE Sens. J.* 20 (4) (2020) 2023–2033.
- [16] C. Tong, A. Palazoglu, X.F. Yan, An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding, *J. Process Control* 23 (10) (2013) 1497–1507.
- [17] X. Wang, U. Kruger, G.W. Irwin, G. McCullough, N. McDowell, Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis, *IEEE Trans. Control Syst. Technol.* 16 (1) (2007) 122–129.
- [18] Z.C. Li, X.F. Yan, Ensemble learning model based on selected diverse principal component analysis models for process monitoring, *J. Chemom.* 32 (4) (2018) e3010.
- [19] J.F. MacGregor, C. Jaeckle, C. Kiparissides, M. Koutoudi, Process monitoring and diagnosis by multiblock PLS methods, *AIChE J.* 40 (5) (1994) 826–838.
- [20] S.J. Qin, S. Valle, M.J. Piovoso, On unifying multiblock analysis with application to decentralized process monitoring, *J. Chemom.* 15 (9) (2001) 715–742.
- [21] S.J. Zhao, J. Zhang, Y.M. Xu, Monitoring of processes with multiple operating modes through multiple principle component analysis models, *Ind. Eng. Chem. Res.* 43 (22) (2004) 7025–7035.
- [22] Q.C. Jiang, X.F. Yan, Monitoring multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability, and Bayesian inference, *Chemometrics and Intelligent Laboratory Syst.* 136 (2014) 121–137.
- [23] Z.Q. Ge, Z.H. Song, Mixture Bayesian regularization method of PPCA for multimode process monitoring, *AIChE J.* 56 (11) (2010) 2838–2849.
- [24] J. Yu, S.J. Qin, Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models, *AIChE J.* 54 (7) (2008) 1811–1829.
- [25] J. Yu, Fault detection using principal components-based gaussian mixture model for semiconductor manufacturing processes, *IEEE Trans. Semicond. Manuf.* 24 (3) (2011) 432–444.
- [26] J.L. Zhu, Z.Q. Ge, Z.H. Song, Distributed Gaussian mixture model for monitoring plant-wide processes with multiple operating modes, *IFAC J. Syst. Control* 6 (2018) 1–15.
- [27] Q.C. Jiang, X.F. Yan, Multimode process monitoring using variational bayesian inference and canonical correlation analysis, *IEEE Trans. Autom. Sci. Eng.* 16 (4) (2019) 1814–1824.
- [28] Q.P. He, J. Wang, Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes, *IEEE Trans. Semicond. Manuf.* 20 (4) (2007) 345–354.
- [29] Q.P. He, J. Wang, Principal component based k-nearest-neighbor rule for semiconductor process fault detection, American Control Conference, IEEE, Seattle, WA, USA, (2008) 1606–1611.
- [30] H.H. Ma, Y. H. H.B. Shi, A novel local neighborhood standardization strategy and its application in fault detection of multimode processes, *Chemometr. Intell. Lab. Syst.* 118 (2012) 287–300.
- [31] X.G. Deng, X.M. Tian, Multimode process fault detection using local neighborhood similarity analysis, *Chin. J. Chem. Eng.* 22 (11) (2014) 1260–1267.
- [32] Y. Zheng, S.M. Mao, S.J. Liu, D.S.H. Wong, Y.W. Wang, Normalized relative RBC based minimum risk Bayesian decision approach for fault diagnosis of industrial process, *IEEE Trans. Ind. Electron.* 63 (12) (2016) 7723–7732.
- [33] Y. Zheng, B. Ai, D.S.H. Wong, S.S. Jang, Y.W. Wang, J. Zhang, An EWMA algorithm with a cycled resetting (CR) discount factor for drift and fault of high-mix Run-To-Run Control, *IEEE Trans. Ind. Informatics* 6 (2) (2010) 229–242.
- [34] Z.Q. Ge, Z.H. Song, Multimode process monitoring based on Bayesian method, *J. Chemom.* 23 (12) (2010) 636–650.
- [35] J.J. Downs, E.F. Vogel, A plant-wide industrial process control problem, *Comput. Chem. Eng.* 17 (3) (1993) 245–255.