

Received February 27, 2020, accepted March 16, 2020, date of publication April 24, 2020, date of current version May 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2989340

# A Novel Monitoring Strategy Combining the Advantages of NPE and GMM

CHENG ZHANG<sup>1</sup>, XUNIAN DAI<sup>1</sup>, XIAOFANG ZHENG, AND YUAN LI

Research Center for Technical Process Fault Diagnosis and Safety, Shenyang University of Chemical Technology, Shenyang 110142, China

Corresponding author: Yuan Li (li-yuan@mail.tsinghua.edu.cn)

This work was supported in part by the Key Program of National Natural Science Foundation of China (NSFC) under Grant 61490701 and Grant 61673279, in part by the Liaoning Natural Science Fund Project under Grant 2019-MS-262, and in part by the Liaoning Provincial Department of Education Fund Project under Grant LJ2019013.

**ABSTRACT** Semiconductor manufacturing process data usually have multimodel and multiphase characteristics which do not meet the application assumptions in neighborhood preserving embedding (NPE). Aiming at the above limitations of NPE, a novel monitoring strategy combining the advantages of the neighborhood preserving embedding and Gaussian mixture model (NPE-GMM) is proposed. Firstly, the window data are obtained by the default window width. Next, the score of the current window data set are calculated by NPE. After that, some Gaussian components of the score are determined by GMM. Finally, a quantification index is proposed to monitor process status. NPE-GMM can not only maintain more local structure information of the current window data set in the feature subspace, but also reduce the computational complexity of GMM in fault detection processes. By introducing the new statistic, NPE-GMM can effectively improve the fault detection rate of some multimodel batch processes. The effectiveness of the proposed method is verified in a numerical case and the semiconductor etching process. The simulation results indicated that the proposed method has a higher fault detection rate than traditional methods.

**INDEX TERMS** Batch production systems, fault detection, Gaussian mixture model, multimodel, moving window, neighborhood preserving embedding, process control, semiconductor manufacturing process.

## I. INTRODUCTION

With the rapid development of science and technology, it has been recognized that the industrial production process is developing towards automation and integration. As an important role, batch processes have been widely used in semiconductor, chemical, fermentation, pharmaceutical and other fields [1], [2]. In order to effectively monitor the status of batch processes, some data-driven methods have been proposed and successfully applied [3]–[5].

As a classical multivariate statistical process control (MSPC) technique, principal component analysis (PCA) has been widely used for dimensionality reduction and process monitoring [6]. Normally, PCA divides the raw space into principal component subspace (PCS) and residual subspace (RS). However, in PCS, the global Euclidean structure of process data is usually preserved, and the local structure information between data points are ignored [7], [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Tedesco<sup>1</sup>.

A method based on local structure, neighborhood preserving embedding (NPE) is a well-applied linear projection technique [9], [10]. In NPE, the raw space is divided into two subspaces by a linear transformation, denoted as feature subspace (FS) and residual subspace (RS). Generally, two statistics, Hotelling's  $T^2$  and squared prediction error (SPE), are used to monitor the status of processes in these subspaces, respectively. To effectively monitor a process using  $T^2$  and SPE, it is essential that assumed that the process data follow a multivariate Gaussian distribution approximately. However, the unique characteristics of batch processes, such as nonlinearity in most batch processes, multimodel batch trajectories due to product mix, and process steps with variable durations and multiphase, cause that the process data do not follow a Gaussian distribution [11]–[13]. Hence, the conventional monitoring methods based on PCA and NPE may result in bias on fault detection [14]. In view of the multiway characteristic of batch process data, some special techniques are developed for monitoring purposes [15]. For instance, multiway PCA (MPCA) proposed by Nomikos *et al.* is utilized in

a semibatch reactor for the production of styrene-butadiene latex [16], multiway kernel PCA(MKPCA) proposed by Lee *et al.* is applied in a simulation benchmark of fed-batch penicillin production [17]. It should be noted that these techniques are only suitable for monitoring the status of a single-model process. Their fault detection capability is degraded when the status of a multimodel process is monitored. Considering above the issues, Gaussian mixture model(GMM) is proposed and extensively employed in batch processes [18], [19]. Nevertheless, GMM may not efficiently capture the local features for some complex processes [20].

Above all, a novel monitoring strategy combining the advantages of the neighborhood preserving embedding and Gaussian mixture model(NPE-GMM) is proposed. The paper is organized as follows. In section 2, the conventional NPE, MNPE and GMM are briefly introduced. The proposed monitoring strategy is introduced in section 3. The result of a numerical case is presented in section 4. Section 5 illustrates the performance of NPE-GMM based monitoring method through a simulation of the batch semiconductor etching process. Section 6 gives conclusions of this paper.

## II. BASIC THEORY

### A. MULTIWAY NEIGHBORHOOD PRESERVING EMBEDDING

In this section, the conventional NPE and MNPE are briefly introduced. Suppose that  $X = [x(1), x(2), \dots, x(m)]^T$  is a data set with  $m$  samples, where  $x(i) \in R^n$  ( $i = 1, 2, \dots, m$ ).

In NPE, we can determine weight matrix  $W$  by minimizing reconstruction error as follows:

$$\phi(W) = \min \sum_{i=1}^m \left\| x(i) - \sum_{j=1}^m w_{ij}x(j) \right\|^2 \quad s.t. \sum_{j=1}^m w_{ij} = 1 \quad (1)$$

where  $w_{ij}$  is a weight parameter between  $x(i)$  and  $x(j)$ , and if  $x(j)$  is not the first  $k$  neighbors of  $x(i)$ ,  $w_{ij} = 0$ .

Generally, we suppose that  $y=Xa$ . The vector  $a$  is computed through minimizing the following cost function under appropriate constraints [10]. The cost function is as follows:

$$\min \sum_{i=1}^m \left\| y(i) - \sum_{j=1}^m w_{ij}y(j) \right\|^2 \quad s.t. a^T X^T X a = 1 \quad (2)$$

Let  $M=(I-W)^T(I-W)$ , the optimization problem of Eq.(2) can be transformed into a generalized eigenvector problem as follows:

$$X^T M X a = \lambda X^T X a \quad (3)$$

The eigenvalues obtained from Eq.(3) are arranged in ascending order. The eigenvectors corresponding to the first  $d$  eigenvalues constitute the projection matrix  $A = [a_1, a_2, a_3, \dots, a_d]$ .

For batch process data  $X_{I \times J \times K}$  ( $I, J$  and  $K$  denote the number of batches, time steps and process variables

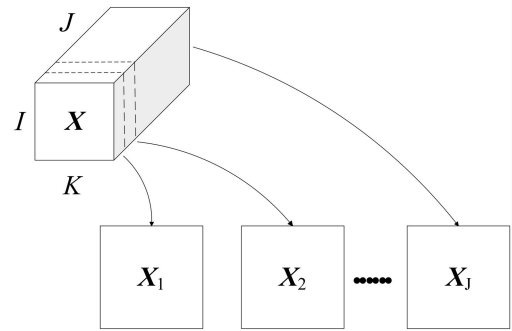


FIGURE 1. Three-dimensional data set unfolded to two-dimensional.

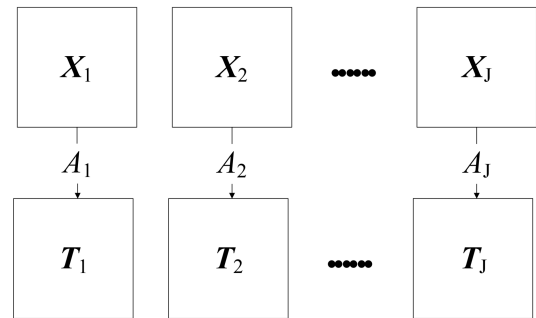


FIGURE 2. The MNPE modeling.

at each time instance, respectively). The basic idea of MNPE is that the conventional NPE is implemented in a two-way matrix unfolded from a raw three-way matrix. In this subsection, the array-data  $X_{I \times J \times K}$  is unfolded into a large matrix  $X_{I \times JK}$ , which is shown in Fig.1. For the  $j$ th time slice matrix  $X_{I \times JK}$  ( $j = 1, 2, \dots, J$ ), its score  $T_{I \times JK}$  as follows:

$$T_{I \times JK} = X_{I \times JK} A_j \quad (4)$$

where  $A_j$  is the projection matrix of conventional NPE, the MNPE modeling is shown in Fig.2.

For monitoring the status of batch process, the  $T^2$  is constructed as the measure of variation in the FS [21]. It is defined as follows:

$$T_{i,j}^2 = x_{i,j} A_j \Lambda_j^{-1} A_j^T x_{i,j}^T \quad (5)$$

where  $x_{i,j}$  represents the  $j$ th time in the  $i$ th batch of  $X$ ,  $A_j$  is calculated by Eq.(3) and  $\Lambda_j$  is the covariance diagonal matrix of matrix  $X_{I \times JK}$ .

### B. GAUSSIAN MIXTURE MODEL

Let  $X = [x(1) x(2) \dots x(m)]^T$  is a data set with  $m$  samples, where  $x(i) \in R^n$  ( $i = 1, 2, \dots, m$ ). GMM assumes that the data consist of  $C$  Gaussian components with different means and variances. The conditional probability  $p(x(i) | C_j)$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, C$ ) that the sample  $x(i)$  belongs to the  $j$ th Gaussian component is calculated

by Eq.(6).

$$p(\mathbf{x}(i) | C_j) = \frac{1}{2\pi |\Sigma_j|^{1/2}} \exp\left(-\frac{(\mathbf{x}(i)^T - \boldsymbol{\mu}(j)^T) \Sigma_j^{-1} (\mathbf{x}(i)^T - \boldsymbol{\mu}(j)^T)^T}{2}\right) \quad (6)$$

where  $\boldsymbol{\mu}(j)$  and  $\Sigma_j$  are mean column vector and covariance matrix of the  $j$ th Gaussian component, respectively, we can calculate the probability that the sample belongs to the  $j$ th Gaussian component as follows:

$$p(\mathbf{x}(i) \in C_j) = \frac{p(\mathbf{x}(i)^T | C_j) P_j}{\sum_{j=1}^C p(\mathbf{x}(i)^T | C_j) P_j} \quad (7)$$

where  $P_j$  represents the prior probability of the  $j$ th Gaussian component.  $\boldsymbol{\mu}_j$ ,  $\Sigma_j$  and  $p(\mathbf{x}(i) | C_j)$  can be iteratively obtained by the EM algorithm.

E-step:

$$P^{(l)}(\mathbf{x}(i) \in C_j) = \frac{P_j^{(l)} p^{(l)}(\mathbf{x}(i)^T | \in C_j)}{\sum_{j=1}^C P_j^{(l)} p^{(l)}(\mathbf{x}(i)^T | \in C_j)} \quad (8)$$

where  $P^{(l)}(\mathbf{x}(i) \in C_j)$  is the posterior probability that the sample belongs to the  $j$ th Gaussian component during the  $l$ th iteration.

M-Step: (9)–(11), as shown at the bottom of this page, where  $\boldsymbol{\mu}_j^{(l+1)}$  and  $\Sigma_j^{(l+1)}$  are the mean and covariance matrix of the  $j$ th Gaussian component, and the  $P_j^{(l+1)}$  is posterior probability during the  $l + 1$ th iteration.

Calculate the label  $c$  of the  $\mathbf{x}(i)$  using Eq.(12)

$$c = \arg \max_{j=1,2,\dots,C} P(\mathbf{x}(i) \in C_j) \quad (12)$$

where  $c$  represents the index of the  $j$ th Gaussian component in which sample  $\mathbf{x}(i)$  is located.

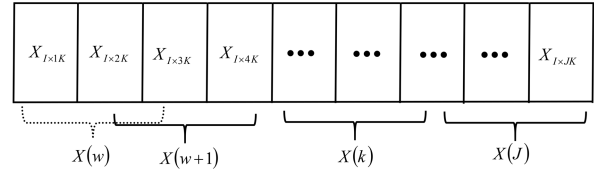


FIGURE 3. Moving window diagram.

### III. A NOVEL MONITORING STRATEGY COMBINING THE ADVANTAGES OF THE NEIGHBORHOOD PRESERVING EMBEDDING AND GAUSSIAN MIXTURE MODEL

#### A. NPE-GMM

Considering the status of samples in multimodel batch processes, a novel monitoring strategy combining the advantages of the neighborhood preserving embedding and Gaussian mixture model(NPE-GMM) is proposed.

In this subsection, NPE-GMM strategy is detailedly introduced. Firstly, a window data are obtained by the default window width. According to the default window width and the moving step, a new window data set is updated through including the newest data and excluding the oldest one [22]. Moving window technique is shown in Fig.3.

Secondly, it is well known that data preprocessing is an important step for making the variables of data  $X(k)$  in the same order of magnitude. The data is preprocessed by the following Eq.(13).

$$\hat{X}(k) = (X(k) - \mathbf{1}_I M(k)) \Lambda^{-1}(k) \quad (13)$$

where  $M(k)$  and  $\Lambda(k)$  are mean column vector and standard deviation diagonal matrix of  $X(k)$ , respectively, and  $\mathbf{1}_I$  is a vector of  $I \times 1$  dimension.

Next, NPE is used to reduce the dimension of the data by Eq.(14), because the moving window technique increases the number of variables dramatically,

$$Y(k) = \hat{X}(k) A_k \quad (14)$$

where  $A_k$  is calculated by Eq.(3).  $Y(k)$  remains most of the local structure information of  $\hat{X}(k)$ .

Then, in  $Y(k)$ ,  $C$  is obtained by cross-validation method [23]. Meanwhile, the  $c$ th Gaussian component

$$\boldsymbol{\mu}_j^{(l+1)} = \frac{\sum_{i=1}^m P^{(l)}(\mathbf{x}(i)^T \in C_j) \mathbf{x}(i)^T}{\sum_{i=1}^m P^{(l)}(C_j | \mathbf{x}(i)^T)} \quad (9)$$

$$\Sigma_j^{(l+1)} = \frac{\sum_{i=1}^m P^{(l)}(\mathbf{x}(i)^T \in C_j) \left(\mathbf{x}(i)^T - (\boldsymbol{\mu}_j^{(l+1)})^T\right)^T \left(\mathbf{x}(i)^T - (\boldsymbol{\mu}_j^{(l+1)})^T\right)}{\sum_{i=1}^m P^{(l)}(\mathbf{x}(i)^T \in C_j)} \quad (10)$$

$$P_j^{(l+1)} = \frac{\sum_{i=1}^m P^{(l)}(\mathbf{x}(i)^T \in C_j)}{m} \quad (11)$$

parameters, such as the mean vector  $\mu_c$  and the diagonal matrix of variance  $\Sigma_c$ , are determined by Eqs.(7-10).

After that, the statistics  $MD_i^2(k)$  of  $Y_i(k)$  ( $i = 1, 2, \dots, I$ ) are calculated by Eq.(15).

$$MD_i^2(k) = \sum_{j=1}^{kk} \left[ \left( Y_i(k) - \mu_c(k) \right) \Sigma_c^{-1/2}(k) - \left( Y_i^j(k) - \mu_c(k) \right) \Sigma_c^{-1/2}(k) \right]^2 \quad (15)$$

$(i = 1, 2, \dots, I)$

where the  $Y_i(k)$  belongs to  $c$ th Gaussian component,  $Y_i^j(k)$  is the  $j$ th neighbor of  $Y_i(k)$  in the Gaussian component. And then Eq.(15) is rearranged as follows:

$$MD_i^2(k) = \sum_{j=1}^{kk} \left[ \left( Y_i(k) - Y_i^j(k) \right) \Sigma_c^{-1/2}(k) \right]^2 \quad (16)$$

Finally, the control limit  $MD_\alpha^2(k)$  is determined according to the kernel density estimation method(KDE) [24]. The confidence level  $\alpha$  in this paper are all selected as 99%.

The detailed fault detection procedure based on NPE-GMM method consists of two parts: offline modeling and online detection.

**Step1: Offline modeling**

For the training sample  $X_{I \times J \times K}$ , the fault detection part consists of seven steps.

- 1). Obtain the  $X_{I \times JK}$  by unfolding  $X$ .
- 2). Determine the width and moving step of the moving window.
- 3). Preprocess of the data  $X(k)$  by Eq.(13).
- 4). Calculate the score  $Y(k)$  by Eq.(14).
- 5). Calculate the parameters of Gaussian components, such as  $C, \mu_c, \Sigma_c$ .
- 6). Calculate the statistic  $MD_i^2(k)$  by Eq.(16).
- 7). Determine the control limit  $MD_\alpha^2(k)$  for fault detection.

**Step2: Online detection**

For the test sample  $\tilde{x}_{J \times K}$ , the fault detection part consists of six steps. The parameters used in the online detection phase are the same as those in the offline modeling phase.

- 1). Obtain the  $\tilde{x}_{1 \times JK}$  by unfolding  $\tilde{x}_{J \times K}$ .
- 2). Preprocess of the data  $\tilde{x}_{J \times K}$  as follows.

$$\hat{x}(k) = (\tilde{x}(k) - M(k)) \Lambda^{-1}(k) \quad (17)$$

- 3). Calculate the score  $y(k)$  by Eq.(14).
- 4). Calculate the label  $c$  of Gaussian component of  $y(k)$  in  $Y(k)$  by Eq.(12).
- 5). Calculate the  $MD^2(k)$  by Eq.(18).

$$MD^2(k) = \sum_{j=1}^{kk} \left[ \left( \tilde{y}(k) - Y^j(k) \right) \Sigma_c^{-1/2}(k) \right]^2 \quad (18)$$

6). Compare  $MD^2(k)$  with the control limits  $MD_\alpha^2(k)$ . If  $MD^2(k) < MD_\alpha^2(k)$ ,  $\tilde{x}_{1 \times JK}$  is classified as a normal sample. Otherwise, it is detected as a fault sample.

**B. NOTICE**

(1). When parameters of the moving window are determined, the moving step is usually chosen as process variables at each time instance  $k$ , and the window width  $w$  is usually set as  $3k$  to  $5k$ , because the  $w$  is too short to effectively obtain the sample status between adjacent moments, and it is too long to make them responsive to process fault.

(2). The dimension of FS is related to moving window width and the dimension of data increases with width.

(3). The parameter  $kk$  in  $MD^2$  should be smaller than the parameter  $k$  in NPE, because NPE keeps the local structure of the sample and its  $k$  nearest neighbors.

(4). The fault detection rate will be different every time GMM is applied, because the results of each iteration of GMM will be slightly different. In order to solve the problem that the fault detection rate is inconsistent in each experiment, the average of the fault detection rate of several experiments in this paper is taken as the final fault detection rate.

**C. ANALYSIS**

In order to reduce the computational complexity, it is necessary to reduce data dimension. Meanwhile, some characteristics of batch process data, such as multimodel and multi-phase, need to be preserved when dimensionality is reduced. As a linear dimensionality reduction algorithm, NPE aims to preserve the local neighborhood structure of data set. Hence, some characteristics of industrial data can be retained in the FS of NPE.

The advantage of GMM is that GMM could monitor process changes without knowing the abnormal pattern [25]. But, in cases for which the data dimension is large, the data are sparse occasionally, so it is difficult to determine GMM model [26]. It is noted that NPE could alleviate these difficulties. Therefore, NPE-GMM is proposed in this paper, which combines the advantages that NPE can contain the local structure of the data in FS and GMM can effectively classify data.

At the same time, NPE-GMM adopts Mahalanobis distance as the statistic which can eliminate the multimodel characteristic of the statistical value. Therefore, compared with some traditional methods, NPE-GMM method has higher fault detection rate in industrial processes with multimodel characteristics.

**IV. CASE STUDY**

In this section, we use an improved dynamic process based on Ref [27] to test the performance of NPE-GMM method in fault detection. The model is given by

$$x(t) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} x(t-1) + \begin{bmatrix} 1.0 & 2.0 \\ 3.0 & -4.0 \end{bmatrix} u(t-1) \quad (19)$$

$$y(t) = x(t) + v(t) \quad (20)$$

TABLE 1. Parameter settings of  $w$ .

		mean	variance
model1	phase 1	0	1
	phase 2	7	0.1
model2	phase 1	-10	4
	phase 2	3	1

TABLE 2. Parameter settings of different fault detection methods.

	$w$	PCs	$k$	$kk$
MPCA	/	2	/	/
MNPE	/	2	9	/
NPE-GMM	2	5	9	6

where  $u$  is the correlated input

$$u(t) = \begin{bmatrix} 0.811 & -0.266 \\ 0.477 & 0.415 \end{bmatrix} u(t-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} w(t-1) \quad (21)$$

where  $u(0) = w(0) = [0 \ 0]^T$ . The normal data have 100 batches and each batch has 1020 sampling instants. The last 1000 samples of each batch are used for analysis because the first 20 samples change greatly. The data vector for modeling consists of  $[y^T(k) \ y^T(k-1) \ u^T(k) \ u^T(k-1)]$ , because input  $u$  and output  $y$  are measured but  $x$  and  $w$  are not.  $v(k)$  is the random noise with zero mean and variance 0.1. These phases are determined by the input  $w$  whose parameter settings are shown in Tab.1

Faults are generated by adding small change in the mean of  $w$  in model 1 after the 501th sample. The 4 normal batches data are selected as the calibration data randomly, and the test batch data are composed of the calibration batch and the fault batch. The plots of monitored variables are shown in Fig.4, where the blue part represents the normal data, while the red line represents the fault data. It can be seen that faults occurred at the 501th time. In this section, we apply MPCA, MNPE and NPE-GMM in this case and then compare their performance. The parameter settings of different methods are listed in Tab.2.

Fault detection results of the calibration data set in MPCA are shown in Fig.5, where there is a significant difference in the statistical values between model 1 and 2. The value of the control limit is completely determined by model 2. At the same time, it can be seen that the statistical values also differ between different stages of the same model, so their control limit centers are different. Its fault detection rate using MPCA are shown in Fig.6. Meanwhile, fault detection rate is 0, because process data with multimodel characteristics do not follow Gaussian distribution. It can be seen that the fault detection rate of MNPE is 0. The main reason that MNPE has the poor performance is the same as MPCA.

The results of GMM classification in FS are shown in Fig.8. It can be found that the data are divided into two models which have different means and covariances. Fault

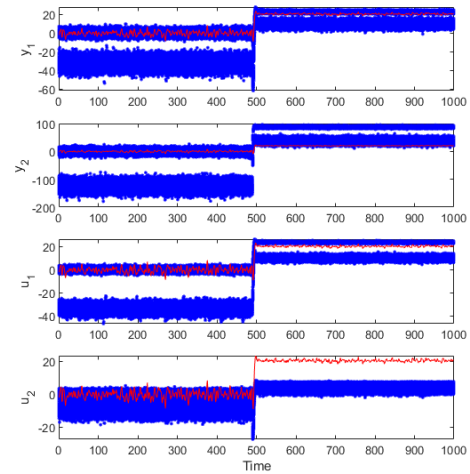


FIGURE 4. The plot of batch data variables.

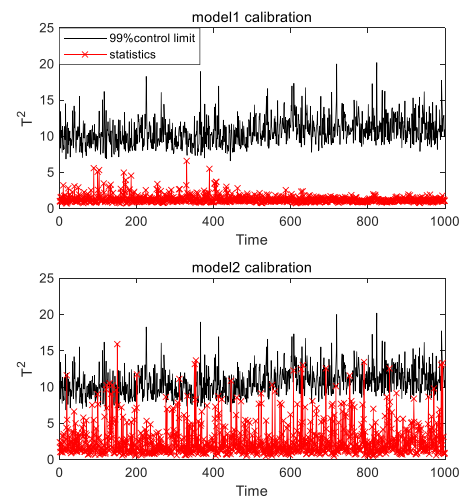


FIGURE 5. Fault detection results of the calibration data set in MPCA.

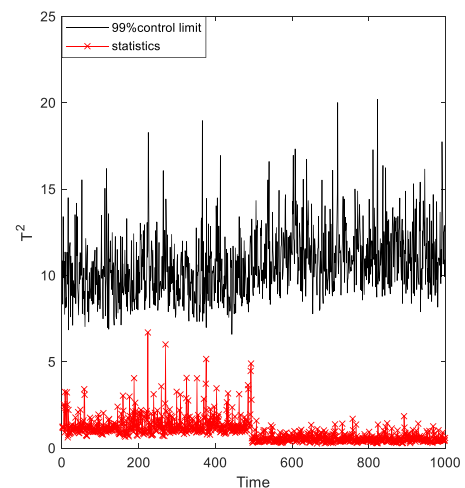


FIGURE 6. Fault detection results using MPCA.

detection results of the calibration data set in NPE-GMM are shown in Fig.9. It can be found that the statistical values and control limits are basically at the same level, which indicates



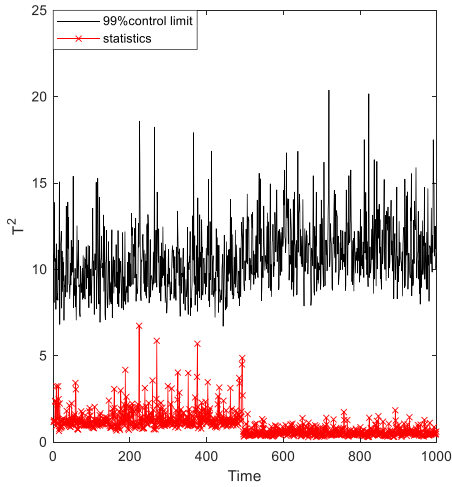


FIGURE 7. Fault detection results using MNPE.

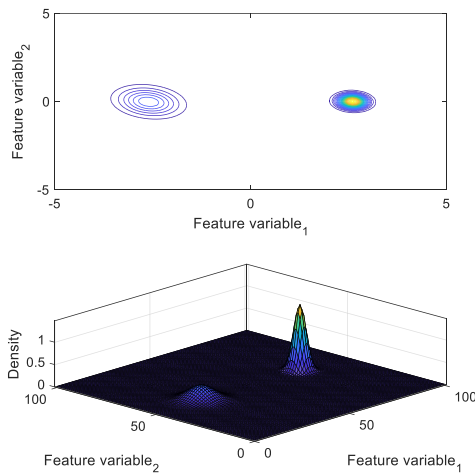


FIGURE 8. The plot of GMM classification results in FS.

that NPE-GMM method eliminates the differences in means and covariances between two models. Fig.10 gives the results of NPE-GMM online fault detection. It can be found that NPE-GMM can timely detects faults. Fig.11 shows the fault detection chromatogram of the test batch, where the black part indicates the fault. We can see that NPE-GMM has a higher fault detection rate and a lower fault alarm rate than other methods. Above all, compared with other methods, we can find that NPE-GMM has the best performance in fault detection.

**V. SEMICONDUCTOR MANUFACTURING PROCESS**

With the rapid development of high-tech industry, semiconductors have been paid more and more attention as an important part of technology products. Therefore, online monitoring of semiconductor production processes has become a research hotspot. In this section, the data set is derived from a semiconductor etching process at Texas instruments [11], [28]. The multiphase is the most important characteristic of this process, which brings some challenges for

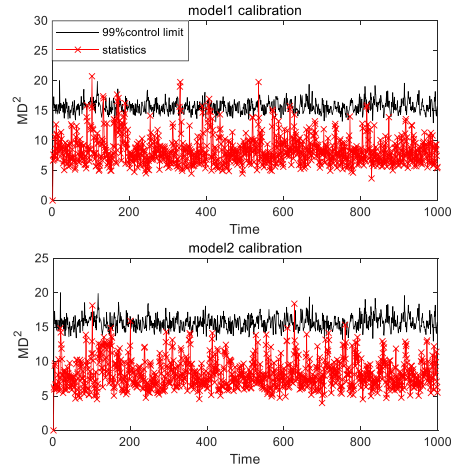


FIGURE 9. Fault detection results of the calibration data set using NPE-GMM.

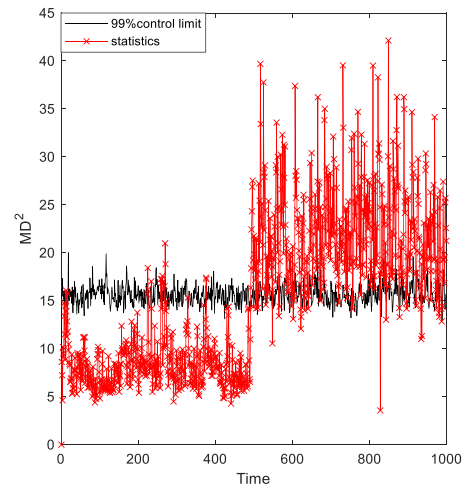


FIGURE 10. Fault detection results using NPE-GMM.

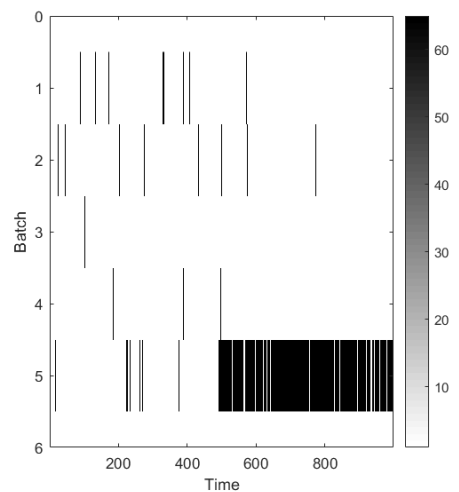
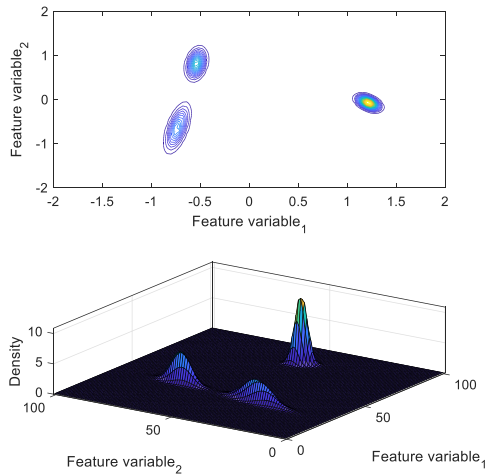


FIGURE 11. The chromatogram of fault detection using NPE-GMM.

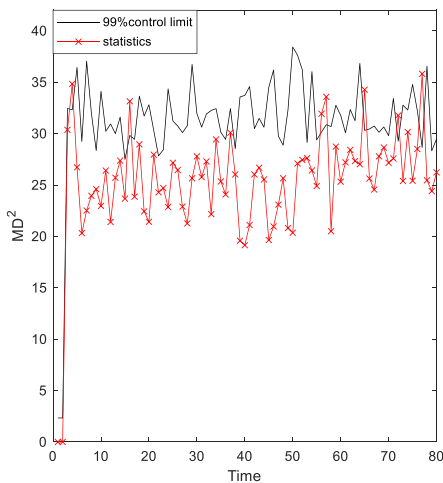
fault detection [4]. The semiconductor data set is generated by three experiments, so data have different means and covariances [10]. The process produced 108 normal wafers and

**TABLE 3. Monitoring variables in the semiconductor etching process.**

No.	Variable Description	No.	Variable Description
1	BCL <sub>3</sub> flow	10	RF power
2	CL <sub>2</sub> flow	11	RF impedance
3	RF bottom A detector	12	TCP tuner
4	Endpoint A detector	13	TCP phase error
5	Helium pressure	14	TCP impedance
6	Chamber pressure	15	TCP top power
7	RF tuner	16	TCP load
8	RF load	17	Vat value
9	Phase error		



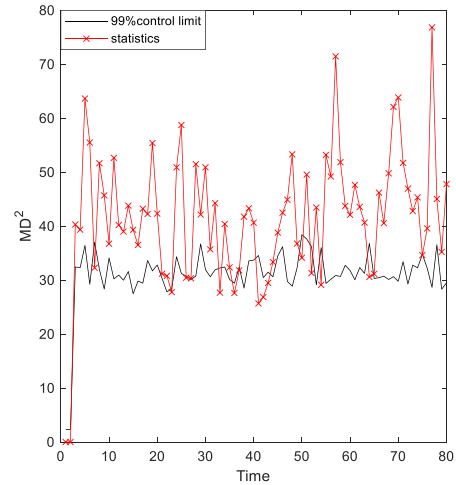
**FIGURE 12. The plot of GMM classification results in FS.**



**FIGURE 13. Fault detection results of the calibration data set using NPE-GMM.**

21 fault wafers. Due to lack of data in some wafers, this paper uses 107 normal wafers and 20 fault wafers for modeling and testing [28]. According to the Ref [29], a total of 17 variable which are shown in Tab.3 are used to monitor the process.

In the data set, different batches have different durations. Therefore, 6-85 sampling instants of semiconductor data are selected for modeling. Test data consist of 3 normal batch data and 20 fault batch data. In this section, a total of 104



**FIGURE 14. Fault detection results of the NPE-GMM in fault.10.**

**TABLE 4. Fault detection rate of different methods.**

Fault No.	MPCA- $T^2$	MNPE- $T^2$	NPE-GMM- $MD^2$
1	13.75	22.5	<b>89.89</b>
2	10	23.75	<b>44.30</b>
3	1.25	1.25	<b>25.36</b>
4	97.5	97.5	<b>100.00</b>
5	1.25	3.75	<b>17.66</b>
6	0	1.25	<b>31.05</b>
7	82.5	96.25	<b>100.00</b>
8	6.25	32.5	<b>55.84</b>
9	1.25	7.5	<b>20.37</b>
10	7.5	17.5	<b>82.19</b>
11	7.5	17.5	<b>42.17</b>
12	<b>100</b>	<b>100</b>	<b>100.00</b>
13	71.25	86.25	<b>100.00</b>
14	8.75	16.25	<b>62.39</b>
15	5	8.75	<b>68.66</b>
16	12.5	36.25	<b>90.17</b>
17	1.25	3.75	<b>69.09</b>
18	47.5	76.25	<b>100.00</b>
19	12.5	48.75	<b>98.29</b>
20	2.5	5	<b>56.84</b>

training batches and 23 test batches are used for modeling and monitoring. By cross-validation [30], the parameter settings as follows: (1).in MPCA, PCs=5; (2) in MNPE, PCs=9,  $k = 9$ ; (3) in NPE-GMM,  $w = 3$ , PCs=20,  $k = 9$ ,  $kk=5$ .

Fig.12 shows the results of GMM classification of the training data in FS. It can be seen that the data are divided into three models with different means and covariances. Therefore, traditional MPCA and MNPE show poor performance in fault detection. Fault detection results of the calibration data set in NPE-GMM are shown in Fig.13, it can be found that the most of normal test samples are below the control limit, which indicates that NPE-GMM model has better stability. In order to illustrate the effectiveness of the proposed method, traditional MPCA and MNPE are tested. Tab.4 gives fault detection results of different methods, it can be found that NPE-GMM has the higher detection rate than others for most of faults. At the same time, the highest fault detection rate of each test batch is marked in bold.

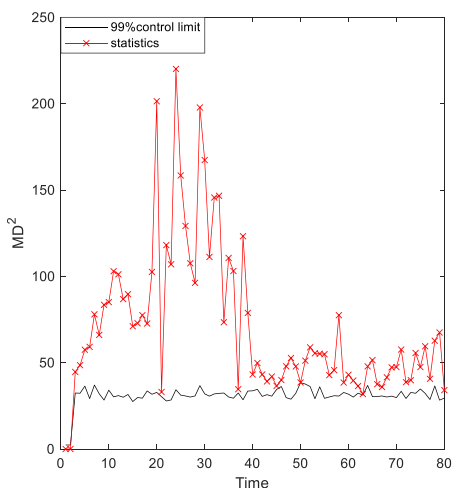


FIGURE 15. Fault detection results of the NPE-GMM in fault19.

Fig.14 and 15 give monitoring results of fault 10 and 19 using NPE-GMM, respectively. From the Ref [31], we know that fault 10 and 19 belong to experiment 30 and 31, respectively. Above all, it can be found that NPE-GMM can effectively identify the faults generated in different models.

## VI. CONCLUSION

In view of the problem that the fault detection in multimodel batch processes exists limitation, a novel monitoring strategy combining the advantages of the neighborhood preserving embedding and Gaussian mixture model is proposed. NPE-GMM can not only effectively eliminate multimodel characteristic of the data through GMM, but also significantly improve fault detection rate by the proposed statistic. The effectiveness of NPE-GMM is verified by a numerical case and the semiconductor batch processes.

## REFERENCES

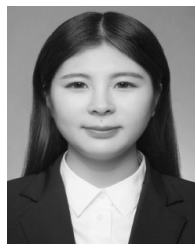
- [1] A. Cinar, *Batch Fermentation: Modeling, Monitoring, and Control*. New York, NY, USA: Marcel Dekker, 2003.
- [2] J.-M. Lee, C. K. Yoo, and I.-B. Lee, "Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis," *J. Biotechnol.*, vol. 110, no. 2, pp. 119–136, May 2004.
- [3] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.
- [4] J. MacGregor and A. Cinar, "Monitoring, fault diagnosis, fault-tolerant control and optimization: Data driven methods," *Comput. Chem. Eng.*, vol. 47, pp. 111–120, Dec. 2012.
- [5] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, Mar. 2013.
- [6] L. Feng, C. Zhang, and Y. Li, "DLNS-PCA-based fault detection for multimode batch process," *CIESC J.*, vol. 69, no. 7, pp. 3159–3166, 2018.
- [7] M. Zhang, Z. Ge, Z. Song, and R. Fu, "Global-local structure analysis model and its application for fault detection and identification," *Ind. Eng. Chem. Res.*, vol. 50, no. 11, pp. 6837–6848, Jun. 2011.
- [8] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *Mach. Learn. Res.*, vol. 4, no. 2, pp. 119–155, 2003.
- [9] A. Miao, Z. Ge, Z. Song, and L. Zhou, "Time neighborhood preserving embedding model and its application for fault detection," *Ind. Eng. Chem. Res.*, vol. 52, no. 38, pp. 13717–13729, Sep. 2013.
- [10] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Beijing, China, Oct. 2005, pp. 1208–1213.
- [11] J. Wang and Q. P. He, "Multivariate statistical process monitoring based on statistics pattern analysis," *Ind. Eng. Chem. Res.*, vol. 49, no. 17, pp. 7858–7869, Sep. 2010.
- [12] C. Zhao, F. Wang, N. Lu, and M. Jia, "Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes," *J. Process Control*, vol. 17, no. 9, pp. 728–741, Oct. 2007.
- [13] Q. P. He and J. Wang, "Fault detection using the k-Nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.
- [14] J. Yu, "A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes," *Chem. Eng. Sci.*, vol. 68, no. 1, pp. 506–519, Jan. 2012.
- [15] T. Chen and J. Zhang, "On-line multivariate statistical monitoring of batch processes using Gaussian mixture model," *Comput. Chem. Eng.*, vol. 34, no. 4, pp. 500–507, Apr. 2010.
- [16] P. Nomikos and J. F. MacGregor, "Monitoring batch processes using multiway principal component analysis," *Amer. Inst. Chem. Eng. J.*, vol. 40, no. 8, pp. 1361–1375, Aug. 1994.
- [17] J.-M. Lee, C. Yoo, and I.-B. Lee, "Fault detection of batch processes using multiway kernel principal component analysis," *Comput. Chem. Eng.*, vol. 28, no. 9, pp. 1837–1847, Aug. 2004.
- [18] J. Yu and S. J. Qin, "Multiway Gaussian mixture model based multiphase batch process monitoring," *Ind. Eng. Chem. Res.*, vol. 48, no. 18, pp. 8585–8594, Sep. 2009.
- [19] J. Yu, "Fault detection using principal components-based Gaussian mixture model for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 24, no. 3, pp. 432–444, Aug. 2011.
- [20] S. Zhang, F. Wang, S. Tan, S. Wang, and Y. Chang, "Novel monitoring strategy combining the advantages of the multiple modeling strategy and Gaussian mixture model for multimode processes," *Ind. Eng. Chem. Res.*, vol. 54, no. 47, pp. 11866–11880, Dec. 2015.
- [21] B. Song, Y. Ma, and H. Shi, "Multimode process monitoring using improved dynamic neighborhood preserving embedding," *Chemometric Intell. Lab. Syst.*, vol. 135, pp. 17–30, Jul. 2014.
- [22] X. Xie and H. Shi, "Dynamic multimode process modeling and monitoring using adaptive Gaussian mixture models," *Ind. Eng. Chem. Res.*, vol. 51, no. 15, pp. 5497–5505, Apr. 2012.
- [23] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models," *AICHE J.*, vol. 54, no. 7, pp. 1811–1829, Jul. 2008.
- [24] Q. Chen, R. J. Wynne, P. Goulding, and D. Sandoz, "The application of principal component analysis and kernel density estimation to enhance process monitoring," *Control Eng. Pract.*, vol. 8, no. 5, pp. 531–543, May 2000.
- [25] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1377–1387, Jul. 2004.
- [26] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometric Intell. Lab. Syst.*, vol. 30, no. 1, pp. 179–196, Nov. 1995.
- [27] G. A. Cherry and S. J. Qin, "Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis," *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 2, pp. 159–172, May 2006.
- [28] Q. P. He and J. Wang, "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," *AICHE J.*, vol. 57, no. 1, pp. 107–121, Jan. 2011.
- [29] S. Zhang, C. Zhao, and F. Gao, "Incipient fault detection for multiphase batch processes with limited batches," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 1, pp. 103–117, Jan. 2019.
- [30] J. Chen and J. Liu, "Mixture principal component analysis models for process monitoring," *Ind. Eng. Chem. Res.*, vol. 38, no. 4, pp. 1478–1488, Apr. 1999.
- [31] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White, and G. G. Barna, "A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process," *J. Chemometrics*, vol. 13, nos. 3–4, pp. 379–396, May 1999.





**CHENG ZHANG** received the B.S. degree in mathematics and applied mathematics from Northeast Forestry University, Heilongjiang, China, in 2003, and the M.S. degree in mathematics and applied mathematics from the Shenyang University of Chemical Technology, Liaoning, China, in 2010. He is currently pursuing the Ph.D. degree with the Department of Control Science and Engineering, Northeastern University, Liaoning.

He is also an Associate Professor with the Shenyang University of Chemical Technology. His research interests include process modeling, monitoring, fault detection, and classification of batch processes.



**XIAOFANG ZHENG** received the B.S. degree in information and computing science from the Shenyang University of Chemical Technology, Liaoning, China, in 2018, where she is currently pursuing the M.S. degree with the Department of Control Theory and Control Engineering.

Her research interests include fault detection and diagnosis of industrial processes.



**XUNIAN DAI** received the B.S. degree in information and computing science from the Shenyang University of Chemical Technology, Liaoning, China, in 2018, where he is currently pursuing the M.S. degree with the Department of Control Theory and Control Engineering.

His research interests include process data analysis, modeling, and fault diagnosis of industrial processes.



**YUAN LI** received the M.S. degree from the Department of Control Science and Engineering, Shenyang University of Chemical Technology, Liaoning, China, in 2001, and the Ph.D. degree from the Department of Control Science and Engineering, Northeastern University, Liaoning, in 2004.

Her research interests include system identification, fault detection and classification, and data-driven complex process fault diagnosis.

...