

Multimode process monitoring based on robust dictionary learning with application to aluminium electrolysis process

Chunhua Yang^a, Longfei Zhou^b, Keke Huang^a, Hongquan Ji^c, Cheng Long^d,
Xiaofang Chen^a, Yongfang Xie^a

^a*School of Information Science and Engineering, Central South University, Changsha 410083, China*

^b*School of Software Engineering, Central South University, Changsha 410083, China*

^c*College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China*

^d*School of Computer Science and Engineering, Nanyang Technological University, Singapore*

Abstract

In modern process industries, many parameters or states can be acquired with sensors, and these parameters or states often have a close relationship with operation conditions. Unfortunately, the process often operates under different modes, and labels thereof are often unknown. In practice, labeling for sampled data is expensive and time-consuming, so identifying the operation conditions of the industrial process is difficult. In addition, sampled data from the industrial system are always contaminated by outliers or noise. Therefore, a robust process monitoring method for the multimode process is particularly important and challenging. In this paper, a robust dictionary learning method is proposed for processes with multiple unknown modes. Firstly, by taking the sparsity of outliers into account, a robust dictionary learning method is proposed to identify and remove the outliers and noise in the sampled training data. Secondly, an iterative minimization algorithm is designed for solving the dictionary learning optimization program. Thirdly, based on the sparsity of the sparse code, we partition the sparse code into different clusters via spectral clustering method, and then the dictionary is divided into some sub-dictionaries according to the cluster results of sparse code. Lastly, when a new sample is generated, we reconstruct it under different sub-dictionaries, and the smallest dictionary

reconstruction error is calculated as a classifier for process monitoring and fault detection. To evaluate the validity and effectiveness of the proposed monitoring approach, we conduct extensive experiments on a numerical simulation, the continuous stirred tank heater (CSTH) process, and an industrial aluminum electrolysis process, in comparison with several state-of-the-art methods. The experimental results demonstrate that the proposed method is able to provide satisfying monitoring results, and it is also robust to outliers in the sampled training data. It is worth mentioning that the proposed method is an unsupervised learning method, therefore, it is more suitable for the process monitoring of real industrial systems.

Keywords: Multimode process monitoring, Outliers, Unsupervised learning, Robust dictionary learning, Aluminum electrolysis process.

1. Introduction

Process monitoring can be defined as a procedure to determine the operating status of an industrial system. Aiming at reliability and high efficiency, detecting the mode or potential abnormalities of industrial systems is extremely
5 important. Recently, more and more industrial systems are beginning to make full use of the recorded and measured data to monitor the system [1, 2, 3, 4, 5, 6]. Based on the large amount of available data, which include sensor measurements, event-logs as well as records, the data-driven techniques have received considerable attention in the research community of industrial process monitoring and
10 control [7, 8, 9, 10]. For example, in the aluminum electrolysis process, we can record the sensor measurements such as cell current, cell voltage, electrolyte level, cryolite ratio, bath temperature as well as the event-logs such as blanking interval [11, 12]. Based on the high-dimensional correlated process data, many multivariate statistical process monitoring (MSPM) methods can be applied.
15 Among these methods, principal component analysis (PCA) and partial least squares (PLS) are two of the most widely used methods [13, 14, 15]. Other complementary MSPM techniques, such as kernel learning methods [16], inde-

pendent component analysis (ICA) [17, 18, 19, 20], canonical variate analysis (CVA) [21, 22] have been studied to alleviate some limitations of PCA and PLS based methods. However, in practice, the industrial process usually works under different operation modes because of the variations of operation condition, market demands and so on [23]. Therefore, it is necessary and significant to study the related process monitoring method for multimode industrial processes.

For multimode process monitoring, some methods have been proposed, such as mix-PCA, multiple PCA and multiple PLS method [24, 25, 26, 27, 28]. These methods all treat different operating modes separately. More recently, Zhao et.al proposed a novel full-condition monitoring strategy based on cointegration and slow feature analysis method [29]. According to these pioneering works, we can see that the collected samples as well as their labels are significant for multimode processes monitoring. However, collection of labeled data is difficult due to the following two reasons: 1) it is expensive to switch the industrial system from one mode to another so as to merely collect the data and their labels, 2) some kinds of modes, especially the faulty modes, in the industrial process are rare and varied in the industrial system. Therefore, unsupervised process monitoring method for unlabeled data is necessary and important for industrial applications. On the other hand, in the industrial system, the outliers always exist in the recorded and measured data because of instrument failure, formatting errors and so on. According to Ref. 30 and Ref. 31, the outliers in the data can render inaccurate decisions for the industrial processes monitor. Therefore, a reliable multimode process monitoring method is meaningful for the industrial application.

Recently, the dictionary learning method, which was an effective statistical machine learning method, was proposed and has gained tremendous attentions in many fields such as pattern recognition, image processing and computer vision due to its excellent generalization ability compared with the traditional data mining methods [32, 33, 34, 35]. Dictionary learning, which is a particular sparse signal model, aims at learning a set of atoms from the recorded data to approximate the given signals linearly. Ren and Lv proposed a fault detection

scheme based on the dictionary learning method for semiconductor manufacturing processes [36]. Ning, Chen and Zhou proposed a sparse contribution plot method based on the label consistent dictionary learning (LCDL) for the multimode chemical process [37]. Although dictionary learning has achieved inspiring success in fault detection of industrial system, robust dictionary learning method of the unlabeled data for the process monitoring and fault detection is still lacking. To the best of our knowledge, this paper is the first work research on an industrial process monitoring characterized by multimode and with outliers in process data based on the dictionary learning method. Therefore, the contribution of this paper involves the development of a robust dictionary learning method for process monitoring. The proposed method can operate for the unlabeled measurement data, and it also can eliminate the outliers in the training data, thus it is more suitable for multimode industrial systems. Compared with several state-of-the-art methods in the research community of industrial process monitoring, the main contributions of this paper can be summarized as follows.

1) Task-oriented. As the operation condition of the industrial system varied frequently, we propose a multimode process monitoring method, which is efficient for the simulation platform as well as industrial aluminum electrolysis process.

2) Compared with the traditional multivariate statistical method, the proposed dictionary learning method can eliminate the negative effect of outliers. Therefore, the proposed method is more reliable for the process monitoring in industrial system.

3) The proposed method is an unsupervised learning method, so it needs less information for the training data, thus, the proposed method is even more adaptive for the modern industrial requirement.

The rest of the paper is organized as follows. In Section 2, the preliminaries with multimode process monitoring method as well as dictionary learning are reviewed. In Section 3, the robust dictionary learning method for multimode process monitoring is formulated. Two kinds of computational experiments are

80 presented to illustrate the performances of the proposed algorithm in Section 4. The industrial aluminum electrolysis process is presented to demonstrate the efficiency of the proposed method in Section 5. Finally, the conclusion is given in Section 6.

2. Preliminaries

85 In this section, we first briefly review some basic backgrounds about dictionary learning, and then analyze the dictionary learning method for process monitoring.

Dictionary learning has been extensively studied due to its important role in sparse representation and low-rank decomposition. Instead of using manual 90 predefined atoms such as wavelets, curvelets or contourlets, dictionary learning methods are data-adaptive and aim at learning a series of atoms from a data set to linearly approximate a given data [38, 39, 40]. Let X be a set of m dimensional M observation data, i.e., $X = [x_1, x_2, \dots, x_M] \in R^{m \times M}$. Dictionary learning is a task that learns a dictionary with K atoms, which can be used to 95 reconstruct the data X sparsely. This task can be accomplished by solving the following problem:

$$\begin{aligned} \langle D, W \rangle = \arg \min_{D, W} \|X - DW\|_F^2 \\ s.t. \|w_i\|_0 \leq T_1, \forall i \end{aligned} \quad (1)$$

Where, $\|X - DW\|_F^2$ represents the reconstruction error, $\|X\|_F = \sqrt{\sum_{i,j} X_{ij}^2}$, and $D = [d_1, d_2, \dots, d_K] \in R^{m \times K}$ is the learned dictionary that contains K atoms, $\{d_j\}_{j=1}^K$. $W = [w_1, w_2, \dots, w_M] \in R^{K \times M}$ is the sparse code matrix of training 100 data X , and T_1 is a sparsity constraint factor, which means that each column of matrix W has fewer than T_1 nonzero items. The original K-SVD algorithm [41] is an iterative approach to obtain the dictionary D . After obtaining the dictionary D , when a new data x_{new} is coming, it can be represented by solving

the following optimization program:

$$\begin{aligned} \hat{w} &= \arg \min_w \|x_{new} - Dw\|_2^2 \\ s.t. & \|w\|_0 \leq T_1 \end{aligned} \quad (2)$$

Where $\|x\|_0$ is L_0 -norm of vector x , which represents the number of nonzero elements in vector x . As L_0 -norm is non-convex, so the solution of (2) is NP-hard. Approximately, one may use the L_1 -norm optimization to replace the L_0 -norm optimization. That is:

$$\begin{aligned} \hat{w} &= \arg \min_w \|x_{new} - Dw\|_2^2 \\ s.t. & \|w\|_1 \leq T_1 \end{aligned} \quad (3)$$

105 The solution of (3) can be obtained by matching pursuit (MP) algorithm or orthogonal matching pursuit (OMP) algorithm [42]. After obtaining the sparse code \hat{w} , which can be used as a feature for classification of the new data x_{new} directly.

Although the dictionary learning method can be used for classification task, 110 when it is applied to industrial system, it may encounter some difficulties. For example, the training data X may be collected from different unknown modes, and it also may be corrupted by outliers in practical applications. Moreover, these difficulties are extremely common in the industrial system. A simple illustration of the effect of outliers on the process monitoring result is vividly shown 115 in Figure 1. From the subfigure 1 (a), we observe that when the multimode training data has no outliers, the faulty data can be classified correctly. However, from the subfigure 1 (b), we can see that when the multimode training data are corrupted by three outliers, one of the decision boundary is severely affected by the outliers, and faulty data cannot be classified correctly. On the 120 contrary, when the proposed method is applied (as subfigure 1 (c)), the negative effect of outliers can be eliminated, and the decision boundary as well as the faulty data can be identified correctly. Next, we proceed to discuss the robust dictionary learning with application to multimode process monitoring.

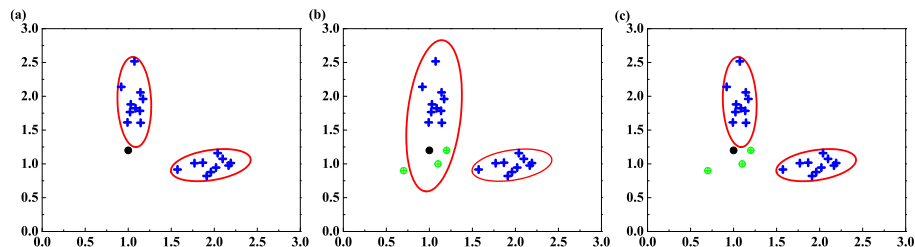


Figure 1: Effect of outliers on the multimode process monitoring. (a) is the original training data without outliers, (b) is the training data with three outliers which are marked with green color, and the monitor result is severely affected by the outliers, (c) is the monitor result of the proposed method. Apparently, the proposed method can eliminate the negative effect of outliers.

3. Multimode Process Monitoring Method

125 In this section, we proposed a novel process monitoring method based on the dictionary learning method. Totally, there are two parts in the proposed method: offline dictionary learning and online multimode process monitoring. In the first part, we model the measurement data in a generative fashion, and then proposed a robust dictionary learning method, which can identify and remove
 130 the outliers in the measurement data. In addition, a revised K-SVD method is proposed for solving the robust dictionary learning objective function. In the second part, the learned dictionary is divided into some sub-dictionaries by spectral clustering, and then when a new measurement is coming, we can represent it by the sub-dictionaries, and the smallest dictionary reconstruction
 135 error (DRE) is used as the classification feature for process monitoring.

3.1. Robust Dictionary Learning

Consider an industrial process, and suppose that we have collected M samples with m sensors, which are denoted by the data matrix $X = [x_1, x_2, \dots, x_M] \in R^{m \times M}$. Most dictionary learning methods [33, 43] make a basic assumption that
 140 the observed data consist of a sparse linear combination of codebooks and some dense Gaussian noises of small variation. Though working well generally, the assumption is unsustainable in case of large corruptions or outliers. In the

industrial process, the data sampled from sensors are always affected by the corruptions or outliers due to the instrument failure, formatting errors and so
 145 on. Inspired by the previous works [39, 43, 44], we assume that each sampled data can be described by the following formation:

$$x_i = x_i^0 + o_i + n_i \quad (4)$$

where, $x_i^0 \in R^m$ is the clean data which represents the measurement without noise and outliers. $n_i \in R^m$ is the measure noise, with the feature of dense, randomness and low amplitude. $o_i \in R^m$ denotes the deterministic outliers in
 150 the measurement data. Typically, as instrument failure or formatting errors occurred only in a small number of sensors, thus o_i is a sparse vector. Namely, only small part of the elements in vector o_i have non-zero values. Moreover, different from measure noise n_i , the non-zero values in vector o_i often have great amplitude.

On the other hand, as the raw materials, static setpoints, operating conditions and market demands are varied, the industrial system frequently works under different operation modes. Intuitively, if two sampled datasets belong to the same class or category, they often lie close in the same subspace. Consequently, the atoms in the dictionary often with the character of discrimination. Therefore, the dictionary learning method can be used to deal with multimode process data with unknown labels. In this paper, a robust dictionary learning method is proposed. First, according to definition of dictionary learning method, the optimization model of the dictionary learning method can be formulated as follows:

$$\begin{aligned} \langle D, W \rangle = \arg \min_{D, W} \|X^0 - DW\|_F^2 \\ s.t. \|w_i\|_0 \leq T_1, \forall i \end{aligned} \quad (5)$$

155 Where, $\|X^0 - DW\|_F^2$ denotes the reconstruction error, $D = [d_1, d_2, \dots, d_K] \in R^{m \times K}$ is the learned dictionary, K is the number of atoms. Specifically, if $K > m$, the dictionary D is called an overcomplete dictionary. $X^0 = [x_1^0, x_2^0, \dots, x_M^0] \in R^{m \times M}$ is the clean data of the industrial process. $W = [w_1, w_2, \dots, w_M] \in R^{K \times M}$

is the sparse code matrix of the training data X^0 . The parameter T_1 is the spar-
160 sity constraint factor, which means that each measure data has fewer than T_1
non-zero elements in the sparse codes. For the model (5), the construction of
dictionary D can be solved by the original K-SVD algorithm [41], which is an
iterative approach to minimize the objective function in model (5) and learns
the sparse code matrix of training data simultaneously. Unfortunately, since
165 the clean data X^0 of the industrial system can hardly be achieved, the previous
method cannot solve the problem exactly [39, 43]. In order to obtain the dic-
tionary and the sparse code matrix, identification and removal the outliers and
noise in the measurement data is necessary. Taking the sparsity of the outliers
 o_i into account, we extend the model (5) by the following formation:

$$\begin{aligned} \langle D, W, O \rangle = \arg \min_{D, W} \|X - DW - O\|_F^2 + \lambda \|O\|_0 \\ \text{s.t. } \|w_i\|_0 \leq T_1, \forall i \end{aligned} \quad (6)$$

170 Where $O = [o_1, o_2, \dots, o_M] \in R^{m \times M}$ denotes the outliers matrix, λ is a positive
constant to control the relative contribution between the reconstruction error
and sparsity of the outliers.

According to the objective function in problem (6), we can see that the
dense noise can be suppressed by the L_F norm of matrix $X - DW - O$, while
175 the sparse outliers can be reduced by the L_0 norm of outliers matrix O . In
order to make optimization program in (6) more easily solvable, the L_0 norm is
approximated by the L_1 norm. So the solution of optimization program in (6)
can be approximated by the solution of the following model:

$$\begin{aligned} \langle D, W, O \rangle = \arg \min_{D, W, O} \|X - DW - O\|_F^2 + \lambda \|O\|_1 \\ \text{s.t. } \|w_i\|_1 \leq T_1, \forall i \end{aligned} \quad (7)$$

where, λ is a positive tuning parameter to balance the reconstruction error and
180 sparsity, which can be determined by many methods, such as grid search [45, 46].
In this paper, we selected it by controlling the value range of T_2 , which equals
the sparsity of outliers matrix O . Although the value of T_1 and T_2 have an effect

on the optimization results, the process monitoring performance is robust for some different combinations of T_1 and T_2 in some given ranges.

185 It is worth noting that the optimization program in (7) is convex with respect to the optimization variables (D, W, O) , hence, it can be solved efficiently and the convergence property of dictionary D can be guaranteed following the analysis of Ref. 41 and Ref. 47. In addition, we should emphasize that the optimization problem (7) cannot be solved directly by the original K-SVD algorithm [41, 43].
190 Empirically, we apply the alternative optimization strategy for the optimization problem (7). In detail, an iterative minimization method is developed to update each variable alternatively. A full description of the method is given in Algorithm 1. In addition, we apply the original K-SVD algorithm to initialize the parameters D^0 , W^0 and O^0 .

Algorithm 1 Iterative minimization method of optimization problem (7).

- 1: **Input:** Training data X .
- 2: **Output:** Dictionary D , Sparse code matrix W .
- 3: **Step 1:** Initialization the dictionary D^0 , sparse code matrix W^0 and the outliers matrix O^0 . Set $J = 1$.
- 4: **Step 2:** Repeat until the stopping criterion is satisfied:
- 5: **Step 2.1 (Sparse coding 1):** Given $D^{(k)}$ and $O^{(k)}$, using the OMP algorithm to update the sparse code matrix $W^{(k)}$ as follow:

$$W^{(k+1)} = \arg \min_W \|X - D^{(k)}W - O^{(k)}\|_F^2$$

$$s.t. \|w_i\|_1 \leq T_1$$

- 6: **Step 2.2 (Sparse coding 2):** Given $D^{(k)}$ and $W^{(k+1)}$, using the OMP algorithm to update the sparse code matrix $O^{(k)}$ as follow:

$$O^{(k+1)} = \arg \min_O \|X - D^{(k)}W^{(k+1)} - O\|_F^2$$

$$s.t. \|o_i\|_1 \leq T_2$$

- 7: **Step 2.3 (Dictionary updating):** Given $W^{(k+1)}$ and $O^{(k+1)}$, using the original K-SVD algorithm to update the dictionary $D^{(k)}$ as follow:

$$D^{(k+1)} = \arg \min_D \|X - DW^{(k+1)} - O^{(k+1)}\|_F^2$$

- 8: **Step 2.4:** Set $J = J + 1$.
-

195 After solving the problem (7), we can obtain the dictionary D , the outliers matrix O and the sparse code matrix W , we will conduct the modes partitioning task. Different from other kinds of modes partitioning methods, such as PCA mixture method [27], which often uses the data matrix X , in this paper, we use the sparse code matrix W to conduct the mode partitioning task. The

200 main reasons are two-fold: 1) data matrix X based methods are sensitive to noise and outliers in the training data, on the contrary, the proposed robust dictionary learning method can eliminate the negative effect of noise and outliers

in the training data and ensure that the dictionary as well as the sparse coding can represent the clean data perfectly, 2) as the label of the training data is
205 unknown, we cannot obtain the modes information from the learned dictionary directly. Fortunately, in the industrial system, data sampled from different operation modes usually show some clustering feature. Namely, if the high-dimensional data belong to the same mode, they often lie in the same low-dimensional subspace, and they can be represented by the similar atoms in
210 the dictionary. Therefore, we should divide the dictionary D into some sub-dictionaries based on the clustering result of sparse code matrix W because of it usually includes fruitful information about the measurement data. In detail, since W is a sparse matrix with block structure, we use the well-known spectral clustering method here [48]. Accordingly, a full description of the dictionary
215 division method is given in Algorithm 2.

Algorithm 2 Dictionary Division Method

- 1: **Input:** Dictionary D , Sparse code matrix W , Cluster number C .
 - 2: **Step 1:** Normalize the columns of W as $w_i \leftarrow \frac{w_i}{\|w_i\|_\infty}$;
 - 3: **Step 2:** Calculate the Euclidean distance matrix ED of normalized sparse code matrix, where $ED_{ij} = \|w_i - w_j\|^2$ and $ED_{ii} = 0$;
 - 4: **Step 3:** Construct a similarity graph with M nodes, where the i_{th} node represents the i_{th} column of ED . The relationship between different nodes is represent by the adjacency matrix Adj , whose elements was calculated by the Gaussian kernel based similarity of ED ;
 - 5: **Step 4:** Calculate the degree matrix De , which is a diagonal matrix with $De_{ii} = \sum_j Adj_{ij}$;
 - 6: **Step 5:** Calculate the Laplace matrix $L = De - Adj$ as well as the normalized Laplace matrix $L_N = I - De^{-1/2} * Adj * De^{-1/2}$;
 - 7: **Step 6:** Calculate the C smallest eigenvalues as well as the corresponding eigenmatrix V of the normalized Laplace matrix L_N ;
 - 8: **Step 7:** Apply k -means clustering method [48] to the eigenmatrix V , and accordingly obtain the segmentation W_1, \dots, W_C ;
 - 9: **Step 8:** For $i = 1 : C$, calculate the main atoms of W_i , and record the i_{th} sub-dictionary by $D_i = [d_{i1}, \dots, d_{iC_i}]$;
 - 10: **Output:** Sub-dictionaries: D_1, \dots, D_C .
-

It is worth noting that the spectral clustering method can automatically optimize the number of clusters by the method of spectrum analysis. To summarize, given the measurement process data, we can obtain the sub-dictionaries, which are corresponding to the different operation modes. Therefore, these sub-dictionaries can be used for classification based on the reconstruction-based scheme. It is worth mentioning that in the whole stage of offline dictionary learning, we only use the unlabeled data, so the offline dictionary learning for the industrial process monitoring is an unsupervised learning method.

3.2. Online multimode process monitoring

In the offline dictionary learning stage, we obtained the sub-dictionaries D_1, \dots, D_C , the atoms of all the sub-dictionaries are selected from the dictionary D . Assuming the dictionary D is acquired through the robust dictionary learning method, and a testing data x_{new} is waiting for identification. First, we represent the testing data x_{new} with the sparse representation model:

$$\begin{aligned} w_{new} &= \arg \min_w \|x_{new} - Dw\|_2^2 \\ s.t. \|w\|_1 &\leq T \end{aligned} \quad (8)$$

225 Then, the sparse code w_{new} can be obtained by MP or OMP algorithm [42].
Next, online process monitoring based on the sparse code is introduced. With the multimode taken into account, the testing data x_{new} belongs to at most one of the modes. Therefore, the testing data x_{new} can be expressed as follows:

$$x_{new} = \tilde{D}\tilde{w} \quad (9)$$

230 Where, $\tilde{D} = [D_1, \dots, D_C, I]$ is the augment dictionary which consists of C sub-dictionaries and an identity matrix $I \in R^{m \times m}$. $\tilde{w} = [w_1, \dots, w_C, f]$. Specifically, if a sample belongs to a subspace which was expanded by a sub-dictionary, then f is close to zero-vector.

From the sparse code w_{new} obtained from model (8), we can obtain the reconstruction error under each mode. Taking inspiration from Ref. 37 and
235 Ref. 38, we adopt the selection operator δ to extract the sparse code of each sub-dictionary. In detail, $\delta_i(w_{new}) = P_i^T w_{new} \in R^m, i = 1, 2, \dots, C$, where $P_i = [p_1^i, p_2^i, \dots, p_{k_i}^i]$, and the p_j^i defined as follow:

$$p_j^i = \underbrace{[0, \dots, 0]_{\sum_{l=1}^{i-1} k_l}}_{\sum_{l=1}^{i-1} k_l} \underbrace{[0, \dots, 0, 1, 0, \dots, 0]_{k_i}}_{k_i} \underbrace{[0, \dots, 0]_{\sum_{l=i+1}^C k_l}}_{\sum_{l=i+1}^C k_l}]^T \quad (10)$$

Therefore, the reconstruction of the testing data x_{new} in the sub-dictionary D_i can be expressed $\hat{x}_{new}^i = D_i \times \delta_i(w_{new}) = DP_i P_i^T w_{new}$. The dictionary

240 reconstruction error in the sub-dictionary D_i is calculated as follows:

$$E_i(x_{new}, D_i) = \|x_{new} - \hat{x}_{new}^i\|_2^2 \quad (11)$$

The reconstruction error $E_i(x_{new}, D_i)$ is compared against the threshold of E_{tr} to determine whether the testing data is normal or not. Namely, the testing data x_{new} belongs to mode c if:

$$DRE < E_{tr} \quad (12)$$

Otherwise, the testing data x_{new} is a fault. Here, E_{tr} can be obtained from the training data based on the kernel density estimation (KDE) method empirically. In this paper, we use the univariate kernel estimator, which is defined as follows:

$$\hat{f}_h(x) = \frac{1}{Mh} \sum_{i=1}^M K\left(\frac{x - E_i}{h}\right) \quad (13)$$

where, x is the data point under consideration, E_i is the reconstruction error of the i_{th} sample in the training data set, h is the bandwidth, and $K(x)$ is the nonnegative kernel function. In this paper, we choose the most commonly used uniform kernel function for $K(x)$ [49]. And the dictionary reconstruction error (DRE) is defined as the smallest reconstruction error on different sub-dictionaries, which is used as the feature for classification. Here, the DRE is defined as follows.

$$\begin{aligned} c &= \arg \min_i E_i(x_{new}, D_i) \\ DRE &= E_c(x_{new}, D_c) \end{aligned} \quad (14)$$

To sum up, the proposed method can identify and remove the outliers and noise in the training data, and the process monitoring task is conducted with clean data, thus it is a robust method. In addition, as the learned dictionary can take the multimode data into account, the proposed method can detect the fault as well as identify the operating mode in the testing data. A full description of the proposed multimode process monitoring method is given in Figure 2.

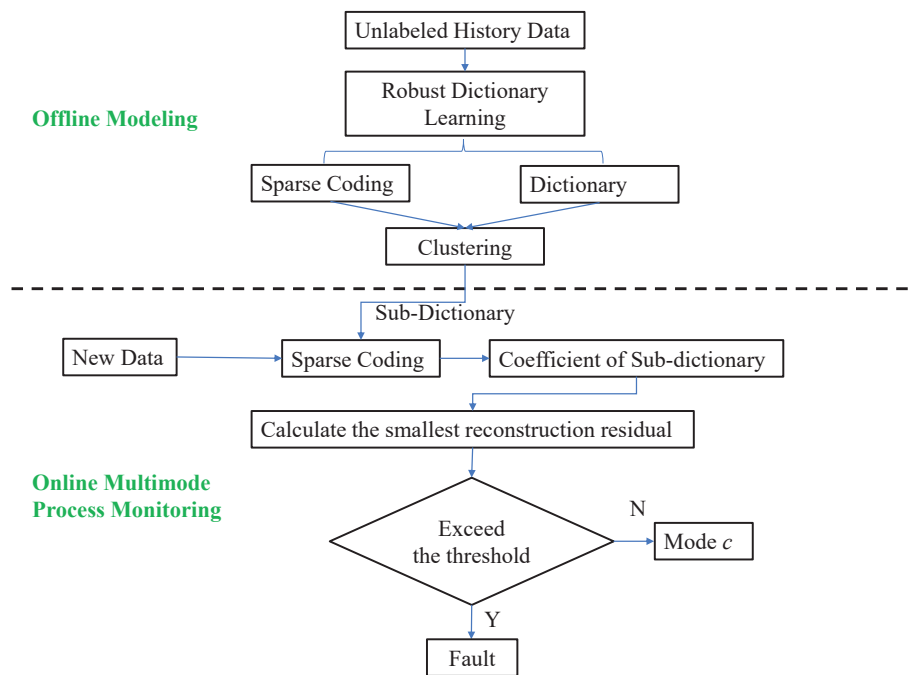


Figure 2: The schematic diagram of the proposed multimode process monitoring method.

260 **4. Illustrative Examples**

In this section, two kinds of examples, a numerical simulation and the CSTH process are carried out to demonstrate the effectiveness of the multimode process monitoring scheme based on the proposed method. To show the effectiveness of the proposed method, we compare it with several state-of-the-art approaches in this field, such as the robust PCA method [50], PCA mixture method [27] and LCDL method [37]. In order to evaluate the performances of different process monitoring methods quantitatively, two generally used indices, i.e. fault detection rate (FDR) and false alarm rate (FAR), are mainly considered [2]. Here, the FDR and FAR are calculated as follow: $FDR = \frac{\#\{J > J_{th} | f \neq 0\}}{\#\{f \neq 0\}} \times 100\%$, $FAR = \frac{\#\{J > J_{th} | f = 0\}}{\#\{f = 0\}} \times 100\%$. $\#\{x\}$ means the number of samples in set x . In addition, in order to simulate the situation when the training data set is corrupted by outliers, some outliers are added into the training data set randomly.

4.1. Numerical simulation

275 First, a multivariate system is selected to verify the performance of the proposed method. Here, the numerical simulation model is introduced in [51] and it can be formulated as:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0.5768 & 0.3766 \\ 0.7382 & 0.0566 \\ 0.8291 & 0.4009 \\ 0.6519 & 0.2070 \\ 0.3972 & 0.8045 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} \quad (15)$$

Where, $x = [x_1, x_2, x_3, x_4, x_5]^T$ is the measurement states used for process monitoring. $s = [s_1, s_2]^T$ represents two independent states, $e = [e_1, e_2, e_3, e_4, e_5]^T$ denotes the observation noise, and all the noises are independent with $e_i \sim N(0, 0.1^2)$, $i = 1, \dots, 5$. For the sake of brevity, we suppose that the linear system works under two different operation modes:

Mode 1: $s_1(k) \sim N(-10, 1), s_2(k) \sim N(-5, 1)$

Mode 2: $s_1(k) \sim U(2, 3), s_2(k) \sim N(7, 1)$

285 Where, $N(\mu, \sigma^2)$ denotes the normal distribution with mean of μ and standard deviation of σ . $U(a, b)$ represents the uniform distribution in the range from a to b .

In the offline dictionary learning stage, 2000 samples are generated from two modes respectively. Hence, there are 4000 samples on the training data. The
290 four methods, namely the robust PCA method, PCA mixture method, LCDL method and the proposed method are all preformed on it. After that, we collected some data as the testing data for the online monitoring stage. In order to verify the performance of the proposed scheme, two kinds of faults: bias fault and multiplicative fault are created as the Table 1. The settings of the different
295 methods are as follow. For robust PCA method, five variables (x_1, x_2, x_3, x_4 and x_5) are used to build the model, and the number of PCs is determined by the cumulative percent variance (CPV) criteria. The regularization parameter of the robust PCA method is set according to its original setting, namely, $\lambda = \frac{1}{\sqrt{\max(m, M)}}, \mu = 10\lambda$. In this experiment, $(\lambda, \mu) = (0.0158, 0.158)$. The
300 threshold of CPV is set as 85%. For PCA mixture method, the mixture components is automatically optimized by the BYY scale-incremental EM algorithm. For LCDL method, the cluster number is 2. For the proposed method, the number of the atoms in the learned dictionary is 40. The ranges of parameters T_1 and T_2 are empirically set as $T_1/m \in [0.01, 1.0]$ and $T_2/m \in [0, 0.5]$, respectively. For all four methods, the thresholds of different indices are determined
305 empirically, and the confidence level is set to $\alpha = 95\%$.

We use the case 1 in the Table 1 to test the generality of the proposed method. The testing data are collected in three steps: 1) the linear system works at Mode 1, and 100 samples are introduced into the testing data, 2) the
310 linear system shifts to the Mode 2, and we collected 100 samples appended to the existing testing data, 3) two bias faults are introduced at the 201st sample. In detail, two bias faults of -1.0 and 0.9 are simultaneously added on x_3 and

Table 1: Two kinds of faults in the numerical simulation

Case Number	Mode	Scenario	Sample Number
Case 1	Mode 1	Normal Operation	1 – 100
	Mode 2	Normal Operation	101 – 200
	Mode 2	Two bias faults of -1.0 and 0.9 are added on x_3 and x_4 respectively	201 – 800
Case 2	Mode 1	Normal Operation	1 – 100
	Mode 2	Normal Operation	101 – 200
	Mode 2	A multiplicative fault is applied to x_2 with a factor of 0.7	201 – 800

x_4 , respectively. The faulty state runs for 600 steps.

In the Figure 3, we show the monitoring results according to the robust PCA method, PCA mixture model, LCDL model and the proposed method. Here, DRR is dictionary reconstruction residual based on the LCDL method, which can be calculated according to Ref. 37. From these subfigures, we can see that the rT^2 and nT^2 of the robust PCA model and PCA mixture method are not able to distinguish the normal data and faulty data exactly. And the $rSPE$ and $nSPE$ of the robust PCA model and PCA mixture method can detect the faulty data, but when apply these statistics to normal data identification, there exist many false alarms. However, as the LCDL method and the proposed method take the multimode characteristic into account, they can identify normal data and faulty data exactly. The quantitative results of FAR and FDR of different methods are listed in Table 2. From the quantitative results, it can be concluded that the proposed method is suitable for the multimode process monitoring.

Next, in order to investigate the robustness of the proposed method for outliers, 200 faulty operation points are randomly added into the training data set. Here, robustness means the method can eliminate the negative effect of

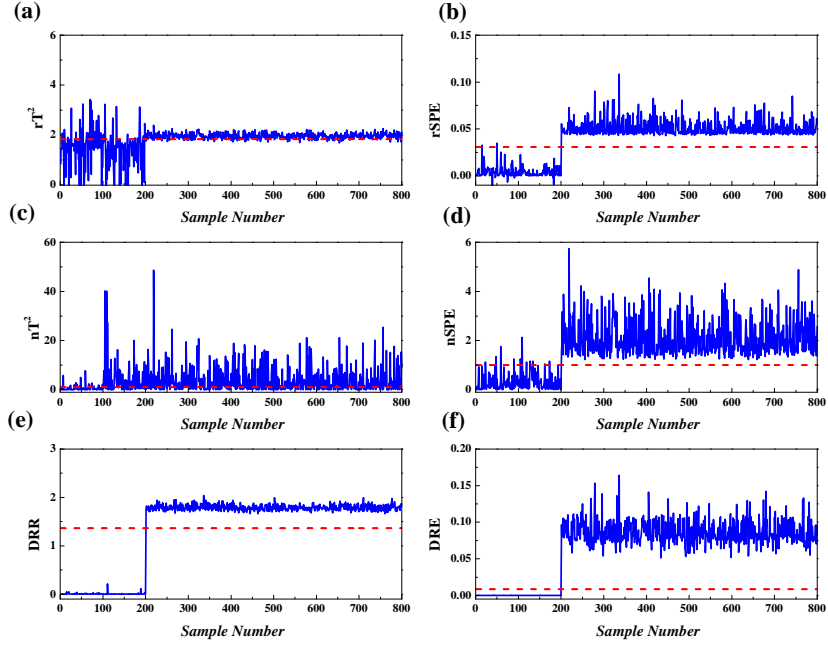


Figure 3: Results of case 1 in the numerical simulation without outliers. Here, (a) is rT^2 statistic of robust PCA model, (b) is $rSPE$ statistic of robust PCA model, (c) is nT^2 statistic of PCA mixture Model, (d) is $nSPE$ statistic of PCA mixture model, (e) is DRR of LCDL method, (f) is DRE of the proposed method.

Table 2: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model in case 1 without outliers

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL Model	Proposed Model
	rT^2	$rSPE$	nT^2	$nSPE$		
FAR	24.00%	1.00%	40.50%	6.00%	0.00%	0.00%
FDR	91.83%	100.00%	60.83%	100.00%	100.00%	100.00%

330 outliers in the training data. The outliers, deviating from the normal oper-
 ation range, are very common in instrument failures and experimental errors
 in real industrial environment. Figure 4 shows the multimode process moni-
 toring results of the robust PCA model, PCA mixture model, LCDL method
 and the proposed method, respectively. From these subfigures, we can see that
 335 the outliers seriously deteriorate the process monitoring results, except for the
 proposed method. In detail, we can see that the nT^2 and $nSPE$ statistics of
 the PCA mixture model are not able to detect the faulty state, the FDR even
 reduce to a very low level. Similarly, the LCDL method is also serious affected
 by the outliers, and the FDR turns to zero. In addition, as the robust PCA
 340 model taken the outliers into account, the $rSPE$ statistics can detect the faulty
 state intermittently, but the FDR is also deteriorated by the outliers. On the
 contrary, as the proposed method taken the sparsity of outliers as well as the
 multimode feature into account, and the process monitoring task was based on
 the reconstructed clean data, thus it can overcome the negative effect on the
 345 outliers and distinguish the normal data and faulty data perfectly.

From the quantitative results in Table 3, we can see that most of these
 methods can identify the normal data correctly. But when the numerical system
 runs in a faulty state, the robust PCA method, PCA mixture method and LCDL
 method fail to detect the faulty state exactly, but the FDR of the proposed
 350 method is still equal to 100%. From the comparison of the four methods, we
 can see that noisy and corrupted training data will reduce the quality of process
 monitoring except for the proposed method. In addition, it is worth noting
 that the proposed method needs less information than other three methods.
 Therefore, we can conclude that the proposed method can eliminate the negative
 355 effect in the training data, and it also can identify the normal data and bias
 fault data of the linear system perfectly, thus it is a robust method for process
 monitoring of linear system with multimode feature. Moreover, we should note
 that the results are steady and only with slightly change for some different
 combinations of dictionary sizes and iteration steps.

360 In order to further demonstrate the robustness of the proposed method for

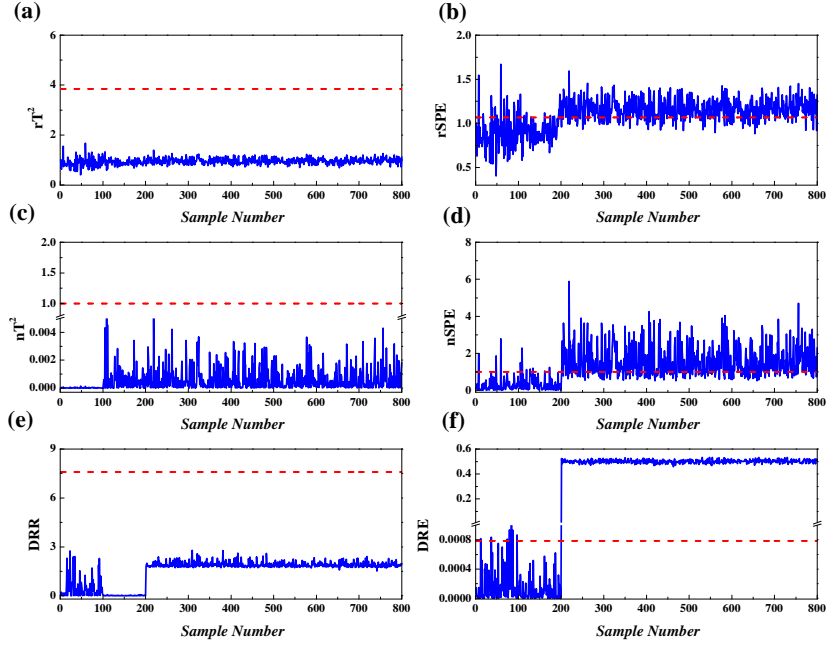


Figure 4: The monitoring results of case 1 with outliers in the numerical simulation. Here, (a) is the rT^2 statistic of robust PCA model, (b) is the $rSPE$ statistic of robust PCA model, (c) is the nT^2 statistic of PCA mixture Model, (d) is the $nSPE$ statistic of PCA mixture Model, (e) is DRR of the LCDL method, (f) is DRE of the proposed method.

Table 3: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model in case 1 with outliers

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL Model	Proposed Model
	rT^2	$rSPE$	nT^2	$nSPE$		
FAR	0.00%	17.00%	0.00%	5.00%	0.00%	3.50%
FDR	0.00%	78.00%	0.00%	69.50%	0.00%	100.00%

Table 4: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model in case 2 with outliers

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL Model	Proposed Model
	rT^2	$rSPE$	nT^2	$nSPE$		
FAR	16.00%	1.50%	0.00%	4.00%	1.00%	2.50%
FDR	88.00%	86.67%	0.00%	94.00%	0.00%	100.00%

outliers, we use the second kind of fault: multiplicative fault, to test the efficiency of the proposed method. For the testing data, the first 200 samples are collected in the same way as the case 1, but in the 200-800 samples, we applied to x_1 with a factor of 0.7. Figure 5 shows the multimode process monitoring results based on the robust PCA model, PCA mixture model, LCDL method and the proposed method, respectively. Since the training data are collected from multimode, and is also corrupted by outliers, the LCDL method fails to provide a satisfied process monitoring results. In addition, the FDR of robust PCA method and PCA mixture method are not good, which means that it can not identify the faulty data exactly. On the contrary, as the proposed method can eliminate the negative effect of outliers, it can distinguish the normal data and faulty data perfectly. Table 4 offers a quantitative results of FDR and FAR based on robust PCA model, PCA mixture model, LCDL method and the proposed method, respectively. It can be observed that both the FDR and FAR of the proposed method are equally or significantly better than that of other methods. Thus, we can derive the conclusion that outliers included in multimode training data set have a great negative effect on FDR and FAR in robust PCA model, PCA mixture method and LCDL method, but the proposed method can deal with the process monitoring characterized by multimode and with outliers in process data. Therefore, we can conclude that the proposed method is a robust method, which can identify the normal data and multiplicative fault data of linear system featured with multimode feature perfectly.

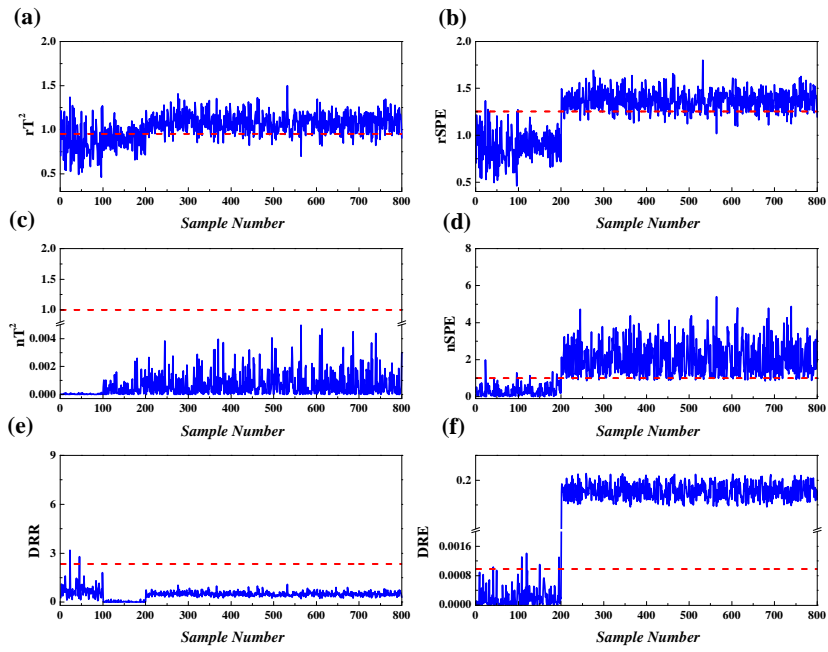


Figure 5: The monitoring results of Case 2 in the numerical simulation. Here, (a) is the rT^2 statistic of robust PCA model, (b) is the $rSPE$ statistic of robust PCA model, (c) is the nT^2 statistic of PCA mixture Model, (d) is the $nSPE$ statistic of PCA mixture Model, (e) is DRR of the LCDL method, (f) is DRE of the proposed method.

4.2. The CSTH process

In this section, we use the CSTH process to testify the efficiency of the proposed multimode process monitoring method. Here, the CSTH process is a realistic platform with nonlinearities, which has been widely used as a benchmark for evaluation of different process monitoring methods [52]. In CSTH process, there exist two physical balances: mass balance and heat balance. Accordingly, the level, flow and temperature measurements are used as the monitoring variable. The schematic diagram of CSTH is given in Figure 6.

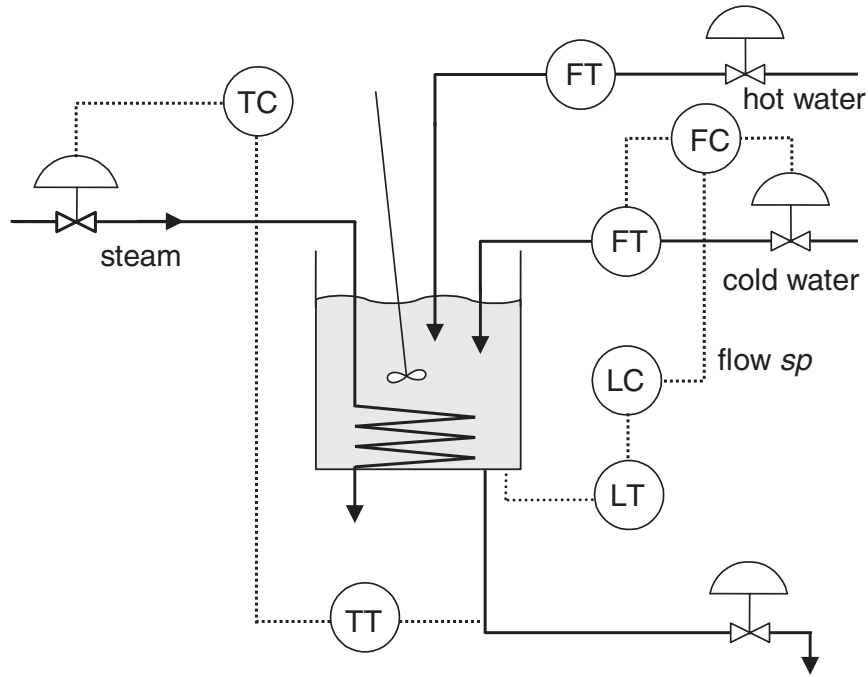


Figure 6: The schematic diagram of CSTH process from [52].

According to the description of CSTH process [52], there exist two operation modes: Mode 1 and Mode 2, which can be implemented by changing the hot water valve from 0 to 5.5. The level setpoint and temperature setpoint of both modes are 12 and 10.5, respectively. The detail information of the two modes are given in Table 5.

Table 5: Two standard operation modes of CSTH process

Mode	Level Setpoint	Temperature Setpoint	Hot water Valve
Mode1	12.0	10.5	0.0
Mode2	12.0	10.5	5.5

In order to demonstrate the effectiveness of the proposed method, we should collect the training data and testing data, respectively. In CSTH process, the data for process monitoring include two parts: the measure data and the output data of controllers. Thus, in the process monitoring, the dimension of training data and testing data is 6. First, 1000 samples are collected from each operation mode as the training data. Here, the training data are used to obtain the learned dictionary via Algorithm 1. Similar to the numerical simulation process, two kinds of testing data are collected following Table 6. The settings of the different methods are as follow. For robust PCA method, six variables are used to build the model, and the number of PCs is determined by the CPV criterion. The regularization parameter of the robust PCA method is set according to its original setting, namely, $(\lambda, \mu) = (0.0224, 0.224)$. The threshold of CPV is set as 85%. For PCA mixture method, the mixture components is automatically optimized by the BYY scale-incremental EM algorithm. For LCDL method, the cluster number is 2. For the proposed method, the number of the atoms in the learned dictionary is 40. The ranges of parameters T_1 and T_2 are empirically set as $T_1/m \in [0.01, 1.0]$ and $T_2/m \in [0, 0.5]$, respectively. For all four methods, the thresholds of different indices are determined empirically, and the confidence level is set to $\alpha = 95\%$.

Similar to the numerical simulation case, we first use the training data without outliers to verify the generality of the proposed method. The testing data are collected according to the Table 6. In the first case, the CSTH process initially works in Mode 1, and 100 samples are collected as the first part of the testing data. Then, the CSTH process shifts to Mode 2, and 100 samples are collected as the second part of the testing data. A bias fault of -0.5 is added to

Table 6: Two kinds of faults in CSTH process

Case Number	Mode	Scenario	Sample Number
Case 1	Mode 1	Normal Operation	1 – 100
	Mode 2	Normal Operation	101 – 200
	Mode 2	A bias faults of -0.5 is added on flow measurements	201 – 800
Case 2	Mode 1	Normal Operation	1 – 100
	Mode 2	Normal Operation	101 – 200
	Mode 2	A multiplicative fault is applied to level with a factor of 0.625	201 – 800

the flow, and run the faulty state for 600 steps. The qualitative and quantitative results of the robust PCA method, PCA mixture method, LCDL method and the proposed method are shown in Figure 7 and Table 7. Clearly, although other three methods can identify the normal data and faulty data with satisfied results, the proposed method is the most accurate method. In addition, due to the labels of training data are draw from cluster method, the LCDL model can not detect the faulty state perfectly. Moreover, as the proposed method taken the multimode into account, it can identify the normal state and faulty state with perfect results. Therefore, we can draw the conclusion that the proposed method is meaningful for the process monitoring in CSTH process.

In order to testify the robustness of the proposed method for outliers, 200 faulty operation points are randomly added into the level data and temperature data. Figure 8 illustrates the process monitoring results of the robust PCA model, PCA mixture model, LCDL method and the proposed method. Due to the training data with feature of multimode as well as outliers, both the robust PCA method, PCA mixture method and LCDL method fail to detect the fault.

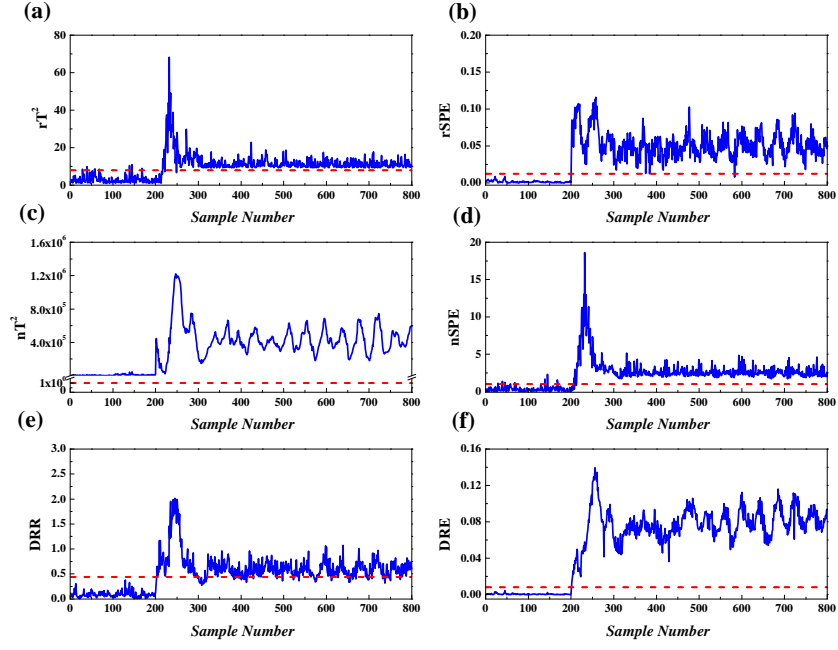


Figure 7: The monitoring results of Case 1 in the CSTH process without outliers. Here, (a) is the rT^2 statistic of robust PCA model, (b) is the $rSPE$ statistic of robust PCA model, (c) is the nT^2 statistic of PCA mixture Model, (d) is the $nSPE$ statistic of PCA mixture Model, (e) is DRR of the LCDL method, (f) is DRE of the proposed method.

Table 7: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model in case 1 of CSTH process without outliers

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL Model	Proposed Model
	rT^2	$rSPE$	nT^2	$nSPE$		
FAR	4.00%	0.00%	100.00%	4.50%	0.00%	0.00%
FDR	97.00%	99.67%	100.00%	98.67%	87.83%	100.00%

However, as the proposed method can identify and remove the outliers in the training data, and conduct the process monitoring based on the reconstructed clean data, thus it can eliminate the negative effects of outliers and identify the normal state and faulty state perfectly. In addition, the monitoring results including the FDR and FAR of robust PCA model, PCA mixture model, LCDL method and the proposed method are listed in Table 8. Comparing with the robust PCA model, PCA mixture model and LCDL method, the proposed method can identify the normal state and faulty state with the best performance in both FDR and FAR indexes.

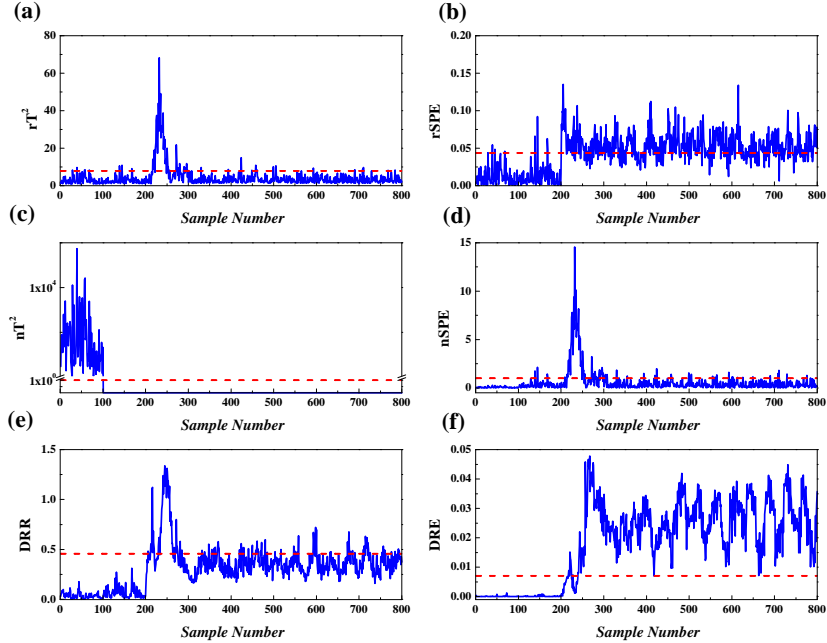


Figure 8: The monitoring results of Case 1 in the CSTH process with outliers. Here, (a) is the rT^2 statistic of robust PCA model, (b) is the $rSPE$ statistic of robust PCA model, (c) is the nT^2 statistic of PCA mixture Model, (d) is the $nSPE$ statistic of PCA mixture Model, (e) is DRR of the LCDL method, (f) is DRE of the proposed method.

In the second case, aiming at verifying the efficiency of the proposed method

Table 8: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model in case 1 of Csth process with outliers

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL	Proposed
	rT^2	$rSPE$	nT^2	$nSPE$	Model	Model
FAR	4.00%	4.00%	50.00%	2.00%	0.00%	0.00%
FDR	11.50%	68.50%	0.00%	14.33%	20.83%	94.50%

on the multiplicative fault of Csth process, we collect the first 200 samples in the same way as case 1, but we impose a multiplicative fault on temperature measurements with a factor of 0.625 for the 200-800 samples of the testing data. Figure 9 illustrates the process monitoring result of the robust PCA model, PCA mixture Model, LCDL method and the proposed method. Since the training data are corrupted by outliers, the robust PCA model and PCA mixture model fails to provide a satisfied process monitoring results, and the LCDL method only detects the fault intermittently. However, as the proposed method can eliminate the negative effects of outliers, it can identify the faulty state from the normal state effectively. In addition, the quantitative monitoring results including the FDR and FAR of the robust PCA model, PCA mixture model, LCDL method and the proposed method are listed in Table 9. Compare with the robust PCA model, PCA mixture model and LCDL method, the proposed method can classify the normal state and faulty state perfectly. Therefore, we can draw the conclusion that the proposed method can detect the additive fault or multiplicative fault of multimode Csth process perfectly, and the process monitoring result is robust even when the training data contaminate by outliers.

5. Aluminum Electrolysis Process Example

High efficiency aluminum electrolysis process has long been a challenging industrial issue. However, due to complex physical and chemical reactions and interference of various external conditions and manual operation, the aluminum electrolysis process often works on different modes, which increase difficulties

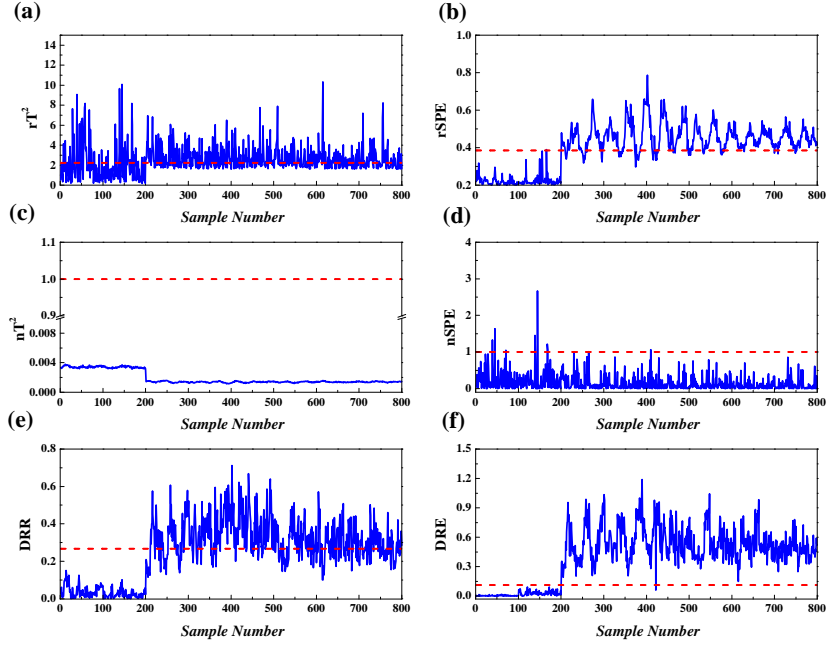


Figure 9: The monitoring results of Case 2 in the CSTH process with outliers. Here, (a) is the rT^2 statistic of robust PCA model, (b) is the $rSPE$ statistic of robust PCA model, (c) is the nT^2 statistic of PCA mixture Model, (d) is the $nSPE$ statistic of PCA mixture Model, (e) is DRR of the LCDL method, (f) is DRE of the proposed method.

Table 9: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model in Case 2 of CSTH process with outliers

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL Model	Proposed Model
	rT^2	$rSPE$	nT^2	$nSPE$		
FAR	31.50%	0.50%	0.00%	3.50%	0.00%	0.00%
FDR	56.17%	87.50%	0.00%	0.33%	67.50%	99.83%

for the aluminum electrolysis process monitoring. In the aluminum electrolysis
470 process, the system must follow two basic balances: material balance and heat
balance. Often, these two balances are hard to measure directly. Alternatively,
the individual anode current is used to characterize the condition of aluminum
electrolysis cells [12]. Differing from cell voltage and cell current, anode current
data can provide an insight into the localized anodic dynamic behavior in an
475 operating aluminum electrolysis cell. In a typical cell, the number of anodes
can vary from 18 to 48, which depends on the cell size. In this paper, the
data set is collected from the 400KA series of the aluminum electrolysis cell of
a aluminum corporation in Shandong province of China. Here, the number of
anodes of the cell is 24, and the schematic diagram of the anodes of a aluminum
480 electrolysis cell is shown in Figure 10. From Figure 10, we can see that the
current flowing through each anode is controlled by the actual cell voltage as
well as the path resistance of each anode. In addition, we also know that
the summation of all currents of the anodes equals the total current supplied
to the cell I_{cell} . The anode effect is an abnormal operation condition closely
485 related to the anode current. When the anode effect occurs, large amounts
of perfluorocarbons such as CF_4 or C_2F_6 are produced, which will decrease
both the current efficiency and the lifetime of the cell. Thus, the spatial anode
currents provide an elaborate perspective for Aluminum Electrolysis process
monitoring. In the following, we will use the 24-dimensional anode current data
490 of two given aluminum electrolysis cells to detect the abnormality of anode
effect.

We selected 6000 samples of daily data as the training data (Here, the first
3000 samples are collected from the first cell, and the last 3000 samples are
collected from the second cell), and selected another 800 samples (The first
495 100 sample are collected from the first cell, and 101th to 200th samples are
collected from the second cell. The last 600 samples are collected from the second
cell when the fault of anode effect happens) as the testing data to verified the
efficiency of the proposed method. The settings of the different methods are as
follow. For the robust PCA method, seven variables are used to build the model,

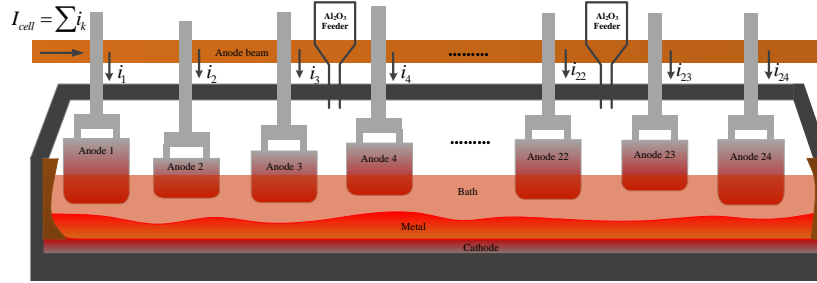


Figure 10: The schematic diagram of the anodes of an aluminum electrolysis cell. Here, the number of anodes in the cell is 24.

500 and the number of PCs is determined by the CPV criterion. The regularization parameter of the robust PCA method is set according to its original setting, namely, $(\lambda, \mu) = (0.0129, 0.129)$. The threshold of CPV is set as 85%. For PCA mixture method, the mixture components is automatically optimized by the BYY scale-incremental EM algorithm. For LCDL method, the cluster number is

505 2. For the proposed method, the number of the atoms in the learned dictionary is 80. The ranges of parameters T_1 and T_2 are empirically set as $T_1/m \in [0.01, 1.0]$ and $T_2/m \in [0, 0.5]$, respectively. For all four methods, the thresholds of different indices are determined empirically, and the confidence level is set to $\alpha = 95\%$.

510 Figure 11 shows the process monitoring results of the robust PCA model, PCA mixture Model, LCDL method and the proposed method. As the data collected from industrial system are often contaminated by different kinds of noise and outliers, we can see that the process monitoring result are all reduced seriously. The rT^2 and nT^2 of robust PCA model and PCA mixture model fail to

515 identify the normal data and faulty data correctly. And the $rSPE$ and $nSPE$ of robust PCA model and PCA mixture model can detect the anode effect correctly but with a high FAR. Compare to the LCDL method, the proposed method can identify the normal and faulty states with high accuracy because it takes the outliers into account. In addition, the quantitative monitoring results including

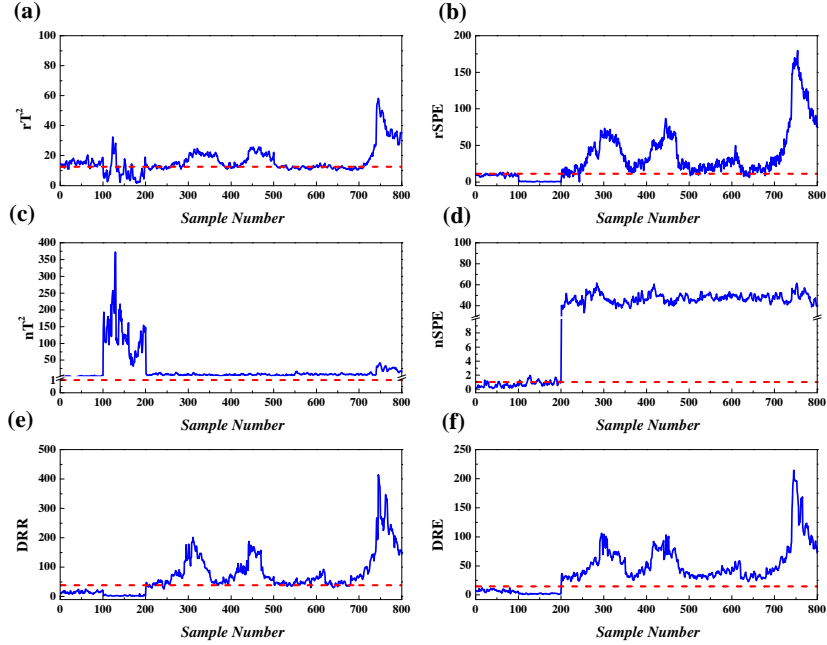


Figure 11: The monitoring results of Aluminum electrolysis process. Here, (a) is the rT^2 statistic of robust PCA model, (b) is the $rSPE$ statistic of robust PCA model, (c) is the nT^2 statistic of PCA mixture Model, (d) is the $nSPE$ statistic of PCA mixture Model, (e) is DRR of the LCDL method, (f) is DRE of the proposed method.

520 the FDR and FAR of robust PCA model, PCA mixture model, LCDL method
and the proposed method are listed in Table 10 . Accordingly, it can concluded
that the proposed method is efficient for industrial process with multimode,
and as it can identify and remove the outliers in the training data, thus it can
eliminate the negative effect caused by the outliers. Therefore, we can conclude
525 that the proposed method is more suitable for the industrial system process
monitoring.

Table 10: Quantitative results of Robust PCA model, PCA mixture Model, LCDL model and the proposed model Aluminium electrolysis process

<i>Index</i>	Robust PCA model		PCA mixture Model		LCDL	Proposed
	rT^2	$rSPE$	nT^2	$nSPE$	Model	Model
FAR	58.00%	6.50%	100.00%	21.50%	0.00%	0.00%
FDR	64.67%	95.33%	100.00%	100.00%	93.50%	100.00%

6. Conclusion and Discussion

In this paper, we introduce a robust dictionary learning method for multi-mode process monitoring. Here, we first model the industrial data in a generative style, and incorporate the feature of sparsity of outliers in the training data, and then identify and remove the outliers based on the revised dictionary learning framework. After that, we conduct the process monitoring task with the reconstructed clean data. Thus, the proposed method can eliminate the negative effect caused by outliers and can be more accurate for process monitoring task. Two kinds of numerical illustrative examples as well as an industrial aluminum electrolysis process example are used to demonstrate the effectiveness of robust multimode process monitoring method. Generally, the proposed dictionary learning method is an unsupervised method. The future work is to take some prior information of the monitored industrial system into the dictionary learning framework, so as to enhance the accuracy of multimode process monitoring further. Although the proposed method can deal with the multimode process monitoring task, it requires all operation conditions be available for model development. Thus, how to deal with the situation when a new operation mode occurs is another interesting direction for future research.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grants No. 61703439, 61803232, 61751312, 61374156), in part

by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 61621062), in part by the 111 Project
550 (B17048), and in part by the innovation project of Central South University (2018zzts613).

References

- [1] S. J. Qin, Survey on data-driven industrial process monitoring and diagnosis, *Ann. Rev. Control* 36 (2) (2012) 220–234.
- 555 [2] S. Yin, S. X. Ding, A. Haghani, H. Hao, P. Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process, *J. Process Control* 22 (9) (2012) 1567–1581.
- [3] S. J. Qin, Process data analytics in the era of big data, *AIChE J.* 60 (9)
560 (2014) 3092–3100.
- [4] H. Ji, X. He, D. Zhou, On the use of reconstruction-based contribution for fault diagnosis, *J. Process Control* 40 (2016) 24–34.
- [5] H. Ji, X. He, J. Shang, D. Zhou, Incipient fault detection with smoothing techniques in statistical process monitoring, *Control Eng. Pract.* 62 (2017)
565 11–21.
- [6] C. Zhao, H. Sun, Dynamic distributed monitoring strategy for large-scale nonstationary processes subject to frequently varying conditions under closed-loop control, *IEEE Transactions on Industrial Electronics* (2018) 1–doi:10.1109/TIE.2018.2864703.
- 570 [7] Z. Ge, Z. Song, F. Gao, Review of recent research on data-based process monitoring, *Ind. Eng. Chem. Res.* 52 (10) (2013) 3543–3562.
- [8] S. Yin, X. Zhu, C. Jing, Fault detection based on a robust one class support vector machine, *Neurocomputing* 145 (2014) 263–268.

- [9] S. Yin, S. X. Ding, X. Xie, H. Luo, A review on basic data-driven approaches for industrial process monitoring, *IEEE Transactions on Industrial Electronics* 61 (11) (2014) 6418–6428.
- [10] S. Yin, X. Zhu, O. Kaynak, Improved pls focused on key-performance-indicator-related fault diagnosis, *IEEE Transactions on Industrial Electronics* 62 (3) (2015) 1651–1658.
- [11] N. A. A. Majid, M. P. Taylor, J. J. Chen, B. R. Young, Multivariate statistical monitoring of the aluminium smelting process, *Comput. Chem. Eng.* 35 (11) (2011) 2457–2468.
- [12] C.-Y. Cheung, C. Menictas, J. Bao, M. Skyllas-Kazacos, B. J. Welch, Characterization of individual anode current signals in aluminum reduction cells, *Ind. Eng. Chem. Res.* 52 (28) (2013) 9632–9644.
- [13] Q. Liu, S. J. Qin, T. Chai, Multiblock concurrent pls for decentralized monitoring of continuous annealing processes, *IEEE Trans. Ind. Electron.* 61 (11) (2014) 6429–6437.
- [14] S. Yin, X. Li, H. Gao, O. Kaynak, Data-based techniques focused on modern industry: An overview, *IEEE Trans. Ind. Electron.* 62 (1) (2015) 657–667.
- [15] H. Ji, X. He, J. Shang, D. Zhou, Incipient sensor fault diagnosis using moving window reconstruction-based contribution, *Ind. Eng. Chem. Res.* 55 (10) (2016) 2746–2759.
- [16] C. Zhao, F. Wang, Y. Zhang, Nonlinear process monitoring based on kernel dissimilarity analysis, *Control Eng. Pract.* 17 (1) (2009) 221–230.
- [17] C. Tong, A. Palazoglu, X. Yan, Improved ica for process monitoring based on ensemble learning and bayesian inference, *Chemom. Intell. Lab. Syst.* 135 (2014) 141–149.

- 600 [18] C. Tong, T. Lan, X. Shi, Double-layer ensemble monitoring of non-gaussian processes using modified independent component analysis, *ISA Trans.* 68 (2017) 181–188.
- [19] C. Tong, T. Lan, X. Shi, Ensemble modified independent component analysis for enhanced non-gaussian process monitoring, *Control Eng. Pract.* 58
605 (2017) 34–41.
- [20] C. Zhao, F. Wang, Z. Mao, N. Lu, M. Jia, Adaptive monitoring based on independent component analysis for multiphase batch processes with limited modeling data, *Industrial & Engineering Chemistry Research* 47 (9) (2008) 3104–3113.
- 610 [21] Z. Chen, S. X. Ding, T. Peng, C. Yang, W. Gui, Fault detection for non-gaussian processes using generalized canonical correlation analysis and randomized algorithms, *IEEE Trans. Ind. Electron.* 65 (2) (2018) 1559–1567.
- [22] Z. Chen, K. Zhang, S. X. Ding, Y. A. Shardt, Z. Hu, Improved canonical correlation analysis-based fault detection methods for industrial processes,
615 *J. Process Control* 41 (2016) 26–34.
- [23] J. Yi, D. Huang, S. Fu, H. He, T. Li, Multi-objective bacterial foraging optimization algorithm based on parallel cell entropy for aluminum electrolysis production process, *IEEE Trans. Ind. Electron.* 63 (4) (2016) 2488–2500.
- [24] S. J. Zhao, J. Zhang, Y. M. Xu, Monitoring of processes with multiple
620 operating modes through multiple principle component analysis models, *Ind. Eng. Chem. Res.* 43 (22) (2004) 7025–7035.
- [25] S. J. Zhao, J. Zhang, Y. M. Xu, Performance monitoring of processes with multiple operating modes through multiple pls models, *J. Process Control* 16 (7) (2006) 763–772.
- 625 [26] C. Zhao, Y. Yao, F. Gao, F. Wang, Statistical analysis and online monitoring for multimode processes with between-mode transitions, *Chemical Engineering Science* 65 (22) (2010) 5961–5975.

- [27] X. Xu, L. Xie, S. Wang, Multi-mode process monitoring method based on pca mixture model, *CIESC Journal* 62 (3) (2011) 743–752.
- 630 [28] C. Zhao, W. Wang, Y. Qin, F. Gao, Comprehensive subspace decomposition with analysis of between-mode relative changes for multimode process monitoring, *Industrial & Engineering Chemistry Research* 54 (12) (2015) 3154–3166.
- [29] C. Zhao, B. Huang, A full-condition monitoring method for nonstationary
635 dynamic chemical processes with cointegration and slow feature analysis, *AIChE Journal* 64 (5) (2018) 1662–1681.
- [30] D. M. Hawkins, *Identification of outliers*, Vol. 11, Springer, 1980.
- [31] H. Ferdowsi, S. Jagannathan, M. Zawodniok, An online outlier identification and removal scheme for improving fault detection performance, *IEEE*
640 *Trans. Neural Netw. Learn. Syst.* 25 (5) (2014) 908–919.
- [32] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (2) (2009) 210–227.
- [33] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration, *IEEE Trans. Image Process.* 17 (1) (2007) 53–69.
645
- [34] M. Elad, M. A. Figueiredo, Y. Ma, On the role of sparse and redundant representations in image processing, *Proc. IEEE* 98 (6) (2010) 972–982.
- [35] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, *Proc. IEEE* 98 (6)
650 (2010) 1031–1044.
- [36] L. Ren, W. Lv, Fault detection via sparse representation for semiconductor manufacturing processes, *IEEE Trans. Semicond. Manuf.* 27 (2) (2014) 252–259.

- [37] C. Ning, M. Chen, D. Zhou, Sparse contribution plot for fault diagnosis of multimodal chemical processes, *IFAC-PapersOnLine* 48 (21) (2015) 619–626.
- [38] D. Wang, S. Kong, A classification-oriented dictionary learning model: Explicitly learning the particularity and commonality across categories, *Pattern Recogn.* 47 (2) (2014) 885–898.
- [39] P. Zhou, C. Fang, Z. Lin, C. Zhang, E. Y. Chang, Dictionary learning with structured noise, *Neurocomputing* 273 (2018) 414–423.
- [40] X. Peng, Y. Tang, W. Du, F. Qian, Multimode process monitoring and fault detection: a sparse modeling and dictionary learning method, *IEEE Trans. Ind. Electron.* 64 (6) (2017) 4866–4875.
- [41] M. Aharon, M. Elad, A. Bruckstein, K-svd: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Proces.* 54 (11) (2006) 4311–4322.
- [42] S. K. Sahoo, A. Makur, Signal recovery from random measurements via extended orthogonal matching pursuit., *IEEE Trans. Signal Proces.* 63 (10) (2015) 2572–2581.
- [43] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1798–1828.
- [44] Z. Xiao, H. Wang, J. Zhou, Robust dynamic process monitoring based on sparse representation preserving embedding, *J. Process Control* 40 (2016) 119–133.
- [45] S. Wang, N. Rahnavard, Binary compressive sensing via sum of l_1 -norm and l_1 (infinity)-norm regularization, in: *Military Communications Conference, MILCOM 2013-2013 IEEE*, IEEE, 2013, pp. 1616–1621.

- 680 [46] Z. Jiang, Z. Lin, L. S. Davis, Label consistent k-svd: Learning a discriminative dictionary for recognition, *IEEE transactions on pattern analysis and machine intelligence* 35 (11) (2013) 2651–2664.
- [47] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, *J. Mach. Learn. Res.* 11 (Jan) (2010) 19–60.
- 685 [48] U. Von Luxburg, A tutorial on spectral clustering, *Stat. Comput.* 17 (4) (2007) 395–416.
- [49] G. R. Terrell, D. W. Scott, Variable kernel density estimation, *The Annals of Statistics* (1992) 1236–1265.
- [50] E. J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, *Journal of the ACM (JACM)* 58 (3) (2011) 11.
- 690 [51] Z. Ge, Z. Song, Multimode process monitoring based on bayesian method, *J. Chemom.* 23 (12) (2009) 636–650.
- [52] N. F. Thornhill, S. C. Patwardhan, S. L. Shah, A continuous stirred tank heater simulation model with applications, *J. Process Control* 18 (3) (2008) 347–360.
- 695