

# Output-Related and -Unrelated Fault Monitoring with an Improvement Prototype Knockoff Filter and Feature Selection Based on Laplacian Eigen Maps and Sparse Regression

Cuiping Xue,\* Tie Zhang, and Dong Xiao



Cite This: *ACS Omega* 2021, 6, 10828–10839



Read Online

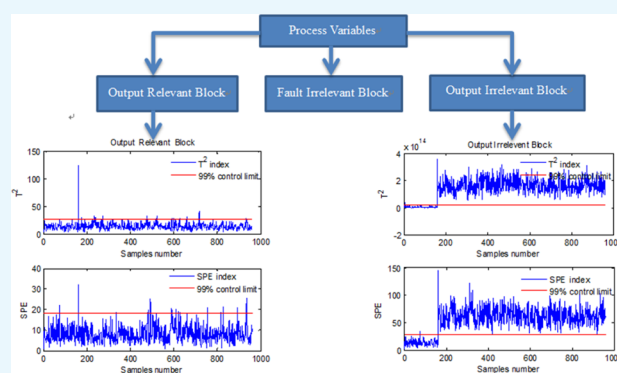
ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** In the process industry, fault monitoring related to output is an important step to ensure product quality and improve economic benefits. In order to distinguish the influence of input variables on the output more accurately, this paper introduces a subalgorithm of fault-unrelated block partition into the prototype knockoff filter (PKF) algorithm for its improvement. The improved PKF algorithm can divide the input data into three blocks: fault-unrelated block, output-related block, and output-unrelated block. Removing the data of fault-unrelated blocks can greatly reduce the difficulty of fault monitoring. This paper proposes a feature selection based on the Laplacian Eigen maps and sparse regression algorithm for output-unrelated blocks. The algorithm has the ability to detect faults caused by variables with small contribution to variance and proves the descent of the algorithm from a theoretical point of view.

The output relation block is monitored by the Broyden–Fletcher–Goldfarb–Shanno method. Finally, the effectiveness of the proposed fault detection method is verified by the recognized Eastman process data in Tennessee.



## 1. INTRODUCTION

Fault monitoring is the key to ensure the long-term stable operation of industrial production processes. Among various fault detection methods, data-driven fault detection in the production process has attracted much attention. Partial least squares (PLS)<sup>1</sup> and principal component analysis (PCA) are widely used. Many algorithms are derived from the basic algorithm.<sup>2–9</sup> Based on the impact on the product output, faults can be divided into output-related faults and output-unrelated faults. Output-related fault means that when some components of input variables deviate from the normal value range for some reason, the values of the corresponding output variables are affected and deviate from the normal value range. On the contrary, output-unrelated fault means that when some components of input variables deviate from the normal value range, the output variables are not affected and remain in the normal value range. There are many output-related fault-monitoring methods. The basic method is to extract the low-dimensional load matrix which can represent the data information from the input data matrix and to make the orthogonal projection of the load matrix to the direction of the output variable to get the output-related information. The problem is that each load vector is a linear combination of all input data. Even though we can monitor the occurrence of faults, it will be very difficult to locate specific variables for the

faults. Therefore, this paper puts forward the idea of dividing the input data into blocks, which is not based on the industrial process but based on the influence of the input data matrix on faults and product output. The prototype knockoff filter (PKF) algorithm<sup>10</sup> divides the input variables into output-related and -unrelated blocks, which is not enough, because when a certain fault occurs, it generally does not affect the values of all input variables. The values of some variables do not fluctuate before and after the fault, which cannot provide any useful information on fault monitoring and brings difficulties to numerical solutions. After confirming the occurrence of the fault, the next fault diagnosis work should determine the position of the fault, and the fewer the variables are selected, the better.

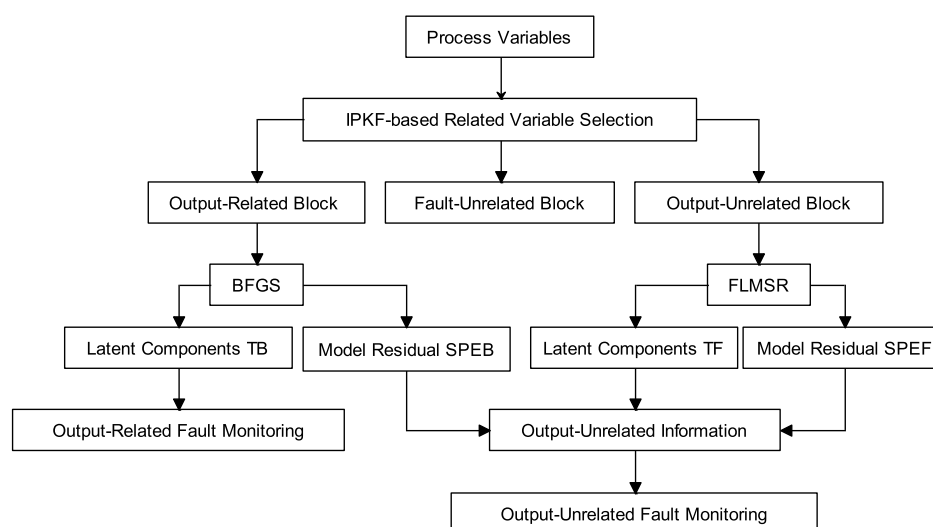
In this paper, an improved PKF (IPKF) algorithm is proposed, which divides the input data matrix into three blocks: fault-unrelated block, output-related block, and output-

Received: January 28, 2021

Accepted: April 6, 2021

Published: April 19, 2021





**Figure 1.** Flowchart of the proposed IPKF-BFGS-FLMSR approach.

unrelated block. The flowchart of the IPKF-BFGS-FLMSR method proposed in this paper is shown in Figure 1.

The PLS method establishes the monitoring model by maximizing covariance between input variables and output variables. By projecting the input and output variables into the corresponding low-dimensional space, the orthogonal eigenvectors of the input and output variables can be obtained. The feature vector emphasizes the interpretation of input variables to output variables, which not only reduces the dimension of variables but also eliminates useless noise. From this feature, it is easy to see that PLS has a weak ability to detect the faults which are unrelated to output. When the input data contain orthogonal information, it will have a negative impact on the predicted output. In addition, the PLS algorithm is easily disturbed by singular points, so it is not suitable for nonlinear regression. Here, we want to keep the advantages of the PLS method and improve its shortcomings. The PLS algorithm belongs to the conjugate direction method in the optimization method. The BFGS method is a quasi-Newton algorithm independently proposed by Broyden, Fletcher, Goldfarb, and Shanno in 1970 to solve unconstrained optimization problems. The BFGS method is also a conjugate direction method. The BFGS algorithm not only satisfies the property of the conjugate direction but also is the quasi-Newton algorithm with the best numerical stability so far. The BFGS algorithm not only has the advantages of the conjugate direction method but also has the advantages of the quasi-Newton method: fast convergence and little iteration. The BFGS algorithm has global convergence and superlinear convergence speed. Therefore, the BFGS algorithm is widely used in various research fields, for example, to generate high-fidelity harmonics,<sup>11</sup> to classify files,<sup>12</sup> and to train weights in neural networks.<sup>13</sup> Because the BFGS algorithm has more advantages than the PLS algorithm, this paper uses the BFGS algorithm for the fault detection of output-related blocks to improve the efficiency of fault detection.

The PCA method selects principal components for data reduction according to the order of variance from large to small, and variables with little variance contribution basically do not affect principal components.

Therefore, when faults occur in variables with little variance contribution, it is difficult for PCA to detect faults, and the

detection efficiency will be very low. Based on this phenomenon, an important dimension reduction method which is called feature selection appears. The feature selection method can remove redundant features from original data and process a large number of high-dimensional data in a short time.<sup>14</sup> The advantage of feature selection is that it can identify the representative features of the original data set and make further calculations easier; subspaces can be classified to enhance the robustness of the algorithm to noise<sup>15</sup> and overfitting of calculation can be prevented.<sup>16</sup> The most interest has been arisen in many scholars because of the advantages of feature selection methods. At present, many methods of feature selection have been created.<sup>15–19</sup> In addition, feature selection methods are being widely used, such as biomedical,<sup>20</sup> commodity recommendation and safety monitoring,<sup>21</sup> speech recognition,<sup>22</sup> text mining,<sup>23</sup> and so on. In the meantime, the feature selection method is divided into three categories: supervised,<sup>24</sup> semisupervised,<sup>16,25,26</sup> and unsupervised.<sup>27–29</sup> Among them, the supervised feature selection method is of high accuracy,<sup>30</sup> and the semisupervised feature selection method can be used when the information is incomplete and only a part of the label information is available. The unsupervised feature selection method cannot provide prior information, so it requires large scales of computation. Inspired by the above methods, this paper proposes a new supervised feature selection method for data dimension reduction (FLMSR). In this method, the influence of variables with small variance contribution on principal components is improved by adding weights, such that faults caused by these variables can be monitored in time. A joint framework of feature selection is introduced into the objective function in the method. The framework integrates the local geometric features in the data retained by Laplacian Eigen maps, and the idea of the linear discriminant analysis (LDA) algorithm can increase the distance between data groups, reduce the distance within data groups, and improve the data recognition ability and  $L_{2,1}$ -norm regularization. Furthermore, the updating rules of the algorithm and the proof of its descent are given. The FLMSR model is used to monitor the faults of output-unrelated blocks; simulation experiments show that the algorithm in this paper can overcome the disadvantage that the PCA algorithm has a high false alarm rate for faults caused

by variables with little variance contribution and has stronger fault detection capability.

The main contributions of this paper are as follows:

1. The PKF algorithm is improved, and the idea of dividing variables into three blocks is put forward to reduce the difficulty of fault location.
2. FLMSR is proposed to reduce the dimension of data. The method integrates Laplacian Eigen maps, LDA algorithm, and  $L_{2,1}$ -norm regularization; a brand-new feature selection framework is formed, which not only preserves the geometric structure of data but also improves the recognition ability of data. The updating rules and descent proof of the algorithm are given.
3. The quasi-Newton algorithm BFGS with the best numerical stability is directly applied to fault detection for the first time.

The rest of this paper is as follows. In Section 2, the PKF algorithm is briefly reviewed. The problem description, algorithm, and optimization process are also explained in this section. Section 3 studies the convergence of the algorithm. Section 4 applies the proposed IPKF-BFGS-FLMSR method to the operation evaluation of TEP and verifies the effectiveness of the proposed method. The fifth part gives the conclusion and outlines the future research work.

## 2. PROPOSED METHODOLOGY

**2.1. Review.** The knockoff filter algorithm is quoted in order to divide the input data into disjoint blocks and detect faults, respectively, in this paper. The knockoff filter is a method where a variable is selected by controlling the error discovery rate.<sup>31</sup> After the proposal, furthermore, the author applied it to high-dimensional data.<sup>32</sup> In the same year, sparse regression and marginal testing were used to create cluster prototypes.<sup>33</sup> Then, the knockoff filter for grouping selection was proposed.<sup>10</sup> The basic idea of the knockoff filter is to construct a knockoff matrix with the same dimension and a covariance matrix as the original data matrix according to some rules, and then, a large augmented matrix is formed together with the original matrix. The output variables are subjected to sparse regression on the augmented matrix, and variables with nonzero regression coefficients corresponding to the original data matrix are output-related variables defined in this paper, and the remaining variables are output-unrelated variables. The reason for grouping selection is that the original data columns have strong correlation and cannot fit the requirements of no singularity. Therefore, the original data columns are grouped to ensure strong correlation within groups and weak correlation between groups. Then, group representatives are selected for each group, and the knockoff filter is run on the data matrix composed of group representatives. If the group representatives are output-related, then the whole group is output-related variables, and other variables are output-unrelated variables.

**2.2. Output-Related and -Unrelated Fault Monitoring Based on IPKF-BFGS-FLMSR.** There is no change in the values of some variables before and after the failure, and any useful information cannot be provided for fault detection, but numerical obstacles are brought to fault detection. Therefore, the following algorithm is proposed to select this part of the variables in this paper (see Table 1). Here, we use Pauta Criterion in statistics to judge the fault data.

As for the data with the fault-unrelated variable deleted, variables with an impact on output or without are only

**Table 1. Create Fault-Unrelated Block Subalgorithm**

algorithm: create fault-unrelated block
1. Input normal data $X_0 \in \mathbb{R}^{N \times m}$ and one fault data as $X_1 \in \mathbb{R}^{n \times m}$ , where $N$ , $n$ , and $m$ represent the number of normal samples, fault samples, and variables, respectively. Column mean vector $\mu$ and column standard deviation vector $\sigma$ of normal data $X_0$ .
2. Remember that the number of data whose value is not in the interval $(\mu(i) - 3\sigma(i), \mu(i) + 3\sigma(i))$ ( $i = 1, 2, \dots, m$ ) in the $i$ th column of fault data $X_1$ is $s(i)$ , and all $s(i)$ constitutes a vector $s$ .
3. The vector $s$ component smaller than $k$ ( $k \ll n$ ) is defined as a fault-unrelated variable, and the subscript of the fault-unrelated variable is put into the fault-unrelated block $C$ .

concerned. If the remaining variables are divided into output-related and output-unrelated parts, the subsequent fault location would be easier. Combined with the previous create fault-unrelated block subalgorithm and the knockoff filter algorithm, the following IPKF algorithm is given (see Table 2),

**Table 2. IPKF Algorithm**

algorithm: improvement prototype knockoff filters
1. Input normal data $X_0$ and fault data $X_1$ .
2. Run the create fault-unrelated block subalgorithm to get invariant variable subscript block $C_3$ and the remaining variable subscript block $C$ ;
3. Run the PKF algorithm on the fault data corresponding to block $C$ ;
4. Blocks $C_1$ and $C_2$ are determined by the regression coefficients obtained in the third step. $C_1$ is the subscript block of output-related variable components, and $C_2$ is the subscript block of output-unrelated variable components.

which can divide the input variables corresponding to the original data into three disjoint blocks, which are fault-unrelated block, output-related block, and output-unrelated block.

Because of the different types of faults and ways of dividing blocks, the following fault detection model is given for the  $i$ th fault (see Table 3).

Here, the calculation of statistics  $T^2$  and SPE used in fault detection is given by the following formula

$$T^2 = \mathbf{t}^T \left( \frac{\mathbf{T}^T \mathbf{T}}{n-1} \right)^{-1} \mathbf{t} \quad (1)$$

$$\text{SPE} = \mathbf{X}^T (\mathbf{I} - \mathbf{P}\mathbf{P}^T) \mathbf{X} \quad (2)$$

When a significance level of  $\alpha$  is given, the threshold value of statistical information is given by the following formula

$$J_{\text{th}, T^2} = \frac{l(n^2 - 1)}{n(n-1)} F_{l, n-l, \alpha} \quad (3)$$

Here,  $n$  is the number of sampled samples and  $l$  is the number of remaining principal components.

$$J_{\text{th}, \text{SPE}} = g \mathcal{X}_{h, \alpha}^2 \quad (4)$$

Here,  $g = \rho^2 / 2\mu$ ,  $h = 2\mu^2 / \rho^2$  and  $\mu$  and  $\rho^2$  are the mean and variance of the statistics of the sample, respectively.

Next, aiming at the existing fault data block, run the offline submodel, collect the way of block partition, load matrix and fault control line, and establish a general fault database following steps are given for online monitoring (see Table 4).

During online monitoring, create a fault-unrelated block for the newly generated sample points and compare it with the fault-unrelated blocks in the fault library. Determine the fault type to which it belongs. Furthermore, determine whether the

Table 3. Offline Submodel

model: offline submodel of <i>i</i> th fault
1. Input normal data $X_0$ and fault data as $X_i \in \mathbb{R}^{n \times m}$ of type <i>i</i> fault together as training data, where <i>n</i> and <i>m</i> represent the number of fault samples and variables, respectively.
2. Use prototype knockoff filters for data $X_i'$ after removing invariant variables. The variables are further divided into an output-related part $C_{i1}$ and an output-unrelated part $C_{i2}$ . For convenience, we write them as blocks $C_{i1}$ , $C_{i2}$ , and $C_{i3}$ . The corresponding data blocks of corresponding data $X_0$ and $X_i$ are denoted as $X_{01}$ , $X_{02}$ , $X_{03}$ , $X_{i1}$ , $X_{i2}$ , and $X_{i3}$ .
3. (1) Use the BFGS algorithm on the data block composed of $X_{01}$ and $X_{i1}$ to get $P_{Bi}$ which is the loading matrix of $X_{01}$ and the threshold $J_{th,T^2,Bi}$ , $J_{th,SPE,Bi}$ of the corresponding statistic $T_{Bi}^2$ , $SPE_{Bi}$ .
(2) Use the FLMSR algorithm on the data block composed of $X_{02}$ and $X_{i2}$ to get $P_{Fi}$ which is the loading matrix of $X_{02}$ and the threshold $J_{th,T^2,Fi}$ , $J_{th,SPE,Fi}$ of the corresponding statistic $T_{Fi}^2$ , $SPE_{Fi}$ .

Table 4. Total Offline Model

model: total offline model
Assuming that there are <i>k</i> fault data blocks, establish a fault database: For $i = 1:k$ Execute offline submodel on fault data $X_i$ to obtain data $C_{i1}$ , $C_{i2}$ , $C_{i3}$ , $P_{Bi}$ , $P_{Fi}$ $J_{th,T^2,Bi}$ , $J_{th,SPE,Bi}$ , $J_{th,T^2,Fi}$ , $J_{th,SPE,Fi}$ of recorded in fault database. End

value of its statistics exceeds the control line of the corresponding statistics and judge whether a fault occurs. If the block obtained is different from any known block of fault-unrelated blocks, it means that a new fault occurs, and it is necessary to collect fault data and update the fault database. The following steps are given for online monitoring (see Table 5).

In order to illustrate the effectiveness of the algorithm, the following three evaluation indicators are introduced in the paper.

$$FAR = \frac{\text{no. of false alarms}}{\text{total normal samples}} \tag{5}$$

$$FDR = \frac{\text{no. of effective alarms}}{\text{total faulty samples}} \tag{6}$$

$$SDR = \frac{\text{no. of effective alarms in normal data} + \text{no. of false alarms in the fault data}}{\text{total samples}} \tag{7}$$

Here, FDR represents the monitoring result of fault samples, FAR represents the monitoring result of normal samples, and SDR represents the monitoring result of all samples.

The three evaluation indices are used to evaluate the performance of the algorithm. In the process of running the above algorithm, the BFGS algorithm should be applied to the fault detection of output-related variables, but it cannot be applied to fault detection directly by itself. Therefore, the

Table 5. Online Monitoring Process

model: online monitoring process
1. For each new observation $x^{new}$ , statistics $x_i^{new}$ belongs to subscript block $C_{new}$ of interval $(\mu(i) - 3\sigma(i), \mu(i) + 3\sigma(i))$ ( $i = 1, 2, \dots, m$ ). Compare $C_{new}$ with $C_{i3}$ in the fault data block to judge the fault data type <i>i</i> to which $x^{new}$ belongs.
2. Using the load matrix $P_{Bi}$ and $P_{Fi}$ of the corresponding <i>i</i> th fault, calculate four statistics $t_{Bnew}^2$ , $spe_{Bnew}^e$ , $t_{Fnew}^2$ , $spe_{Fnew}^e$ and then compare them with the corresponding threshold $J_{th,T^2,Bi}$ , $J_{th,SPE,Bi}$ , $J_{th,T^2,Fi}$ , $J_{th,SPE,Fi}$ .
(1) If $t_{Bnew}^2 > J_{th,T^2,Bi}$ then an output-related fault occurs, and the fault location appears in the variables contained in $C_{i1}$ . If $spe_{Bnew} > J_{th,SPE,Bi}$ an output-unrelated fault occurs, and the fault location appears in the variables contained in $C_{i1}$ .
(2) If one of the two inequalities $t_{Fnew}^2 > J_{th,T^2,Fi}$ , $spe_{Fnew}^e > J_{th,SPE,Fi}$ becomes true, then an output-unrelated fault occurs and the fault location appears in the variables contained in $C_{i2}$ .
3. If it is judged in step 1 that some variables of $x^{new}$ are out of the value range, but they are not consistent with the existing fault types, the fault data can be further collected to form a new fault data matrix, and the offline submodel can be executed to update the fault database.

BFGS algorithm is improved and introduced for fault diagnosis in the paper.

The objective function constructed from the input matrix  $X \in \mathbb{R}^{n \times l}$  and the output matrix  $Y \in \mathbb{R}^{n \times 1}$  is as follows.

$$\begin{aligned} \min f(b) &= \frac{1}{2} \|Y - Xb\|^2 \\ &= \frac{1}{2} (Y^T Y + b^T X^T X b - 2Y^T X b) \end{aligned} \tag{8}$$

Gradient of objective function  $f(b)$  is as follows.

$$\nabla f(b) = X^T X b - X^T Y \tag{9}$$

Mark  $g_k = \nabla f(b_k) = X^T X b_k - X^T Y$  (see Table 5). The detailed steps of the BFGS algorithm in fault detection are given in Table 6.

Table 6. BFGS Algorithm for Fault Diagnosis

algorithm: BFGS
1. Point out the initial point $b_0 \in \mathbb{R}^l$ , the initial quasi-Newton matrix $H_0 \in \mathbb{R}^{l \times l}$ , and the termination limit $\epsilon > 0$ , $k := 0$ .
2. If $\ g_k\  \leq \epsilon$ , stop and output the optimal solution $b_k$ .
3. Calculate $d_k = -H_k g_k$ .
4. Calculate $w_k = d_k / \ d_k\ $ , score vector $t_k = X w_k$ , and load vector $p_k = X^T t_k / t_k^T t_k$ , where $t_k$ is the <i>k</i> th column of $T_B$ and $p_k$ is the <i>k</i> th column of $P_B$ ;
5. Find the step factor $\alpha_k$ by linear search and make $b_{k+1} = b_k + \alpha_k d_k$ ;
6. Correct $H_k$ to produce $H_{k+1}$ , $H_{k+1} = \left( E - \frac{s_k y_k^T}{s_k^T y_k} \right) H_k \left( E - \frac{y_k s_k^T}{s_k^T y_k} \right) + \frac{s_k s_k^T}{s_k^T y_k}$ .
Make quasi-Newton condition $H_{k+1} y_k = s_k$ hold, where $y_k = g_{k+1} - g_k$ , $s_k = b_{k+1} - b_k$ ;
4. $k := k + 1$ , return to step 2.

For a newly measured sample data  $x_{new} \in \mathbb{R}^l$ . Calculate the value of new sample point statistics  $t_{B,new}^2 = x_{new}^T P_B \Lambda^{-1} P_B^T x_{new}$ , where  $\Lambda = \frac{T^T T}{n-1} = \frac{1}{n-1} P_B^T X^T X P_B$  and statistics  $spe_{B,new} = \|x_{new} (I - P_B P_B^T)\|$ . Therefore, we can compare the value of

statistics with the corresponding control limit to judge whether there is a fault.

For the unrelated part of the output in the paper, a new FLMSR (feature selection based on Laplacian Eigen maps and sparse regression) method is proposed. A joint framework of feature selection into the objective function is introduced in the method. It combines the local geometric features among Laplacian Eigen maps data and ideas of the LDA algorithm, increases the distance between data groups, reduces the distance within groups, improves the ability of data recognition and  $L_{2,1}$ -norm regularization term, and increases the sparsity of variables. Therefore, the FLMSR algorithm is introduced as follows.

As the known condition that  $\mathbf{X} \in \mathbf{R}^{n \times t}$  is normal data, and  $\mathbf{Y} \in \mathbf{R}^{2n_1 \times t}$  is data synthesized by normal data and fault data according to the ratio of 1:1 and unitized by the mean and variance of normal data, the constraint minimization problem is defined as follows

$$\begin{aligned} \min \sum_{i=1}^n \|\mathbf{C}_n \mathbf{X} \mathbf{P} - \mathbf{C}_n \mathbf{X} \mathbf{W} \mathbf{P}\|_F^2 \\ + \alpha \text{tr}(\mathbf{W}^T \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} \mathbf{W}) + \beta \|\mathbf{W}\|_{2,1}, \\ \text{s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}, \mathbf{W} \geq 0, \mathbf{P}^T \mathbf{P} = \mathbf{I} \end{aligned} \quad (10)$$

Here,  $\mathbf{C}_n = \mathbf{I}_n - (1/n) \mathbf{1}_n \mathbf{1}_n^T$ ,  $\mathbf{1}_n = (1, 1, \dots, 1)^T$ . The regular term  $\text{tr}(\mathbf{W}^T \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} \mathbf{W})$  in the objective function can keep the geometrical structure of the data and enhances the discriminative capability. Here,  $\alpha$  is a positive regularization parameter. We introduce the  $L_{2,1}$ -norm regularization  $\beta \|\mathbf{W}\|_{2,1}$  to 10 to enforce the sparsity among rows of the transformation matrix  $\mathbf{W}$ . The Gaussian similarity matrix of sample  $\mathbf{Y}$  is defined as  $\mathbf{S} = \left( \exp\left(-\frac{\|\mathbf{Y}^i - \mathbf{Y}^j\|_2^2}{\sigma^2}\right) \right)_{2n_1 \times 2n_1} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$ , where  $\mathbf{Y}^i$  and  $\mathbf{Y}^j$  are the  $i$ th and  $j$ th row of matrix  $\mathbf{Y}$ , respectively. Then, the intraclass similarity matrix is  $\mathbf{S}_w = \begin{pmatrix} \mathbf{S}_{11} & 0 \\ 0 & \mathbf{S}_{22} \end{pmatrix}$  and the interclass similarity matrix is  $\mathbf{S}_b = \begin{pmatrix} 0 & \mathbf{S}_{12} \\ \mathbf{S}_{21} & 0 \end{pmatrix}$ . It is proposed that

$$\begin{aligned} \mathbf{D}_w &= \left( \sum_{i=1}^{2n_1} \mathbf{S}_w(1, i), \sum_{i=1}^{2n_1} \mathbf{S}_w(2, i), \dots, \sum_{i=1}^{2n_1} \mathbf{S}_w(2n_1, i) \right) \\ \mathbf{D}_b &= \left( \sum_{i=1}^{2n_1} \mathbf{S}_b(1, i), \sum_{i=1}^{2n_1} \mathbf{S}_b(2, i), \dots, \sum_{i=1}^{2n_1} \mathbf{S}_b(2n_1, i) \right), \mathbf{L}_w \\ &= \mathbf{D}_w - \mathbf{S}_w, \mathbf{L}_b = \mathbf{D}_b - \mathbf{S}_b \end{aligned}$$

Using the external penalty function method, we write the augmented objective function of eq 9 as follows

$$\begin{aligned} \min F(\mathbf{P}, \mathbf{W}) &= \|\mathbf{C}_n \mathbf{X} \mathbf{P} - \mathbf{C}_n \mathbf{X} \mathbf{W} \mathbf{P}\|_F^2 \\ &+ \alpha \text{tr}(\mathbf{W}^T \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} \mathbf{W}) + \beta \|\mathbf{W}\|_{2,1} \\ &+ \frac{\mu}{2} (\|\mathbf{W}^T \mathbf{W} - \mathbf{I}\|_F^2 + \|\mathbf{P}^T \mathbf{P} - \mathbf{I}\|_F^2) \\ &+ \text{tr}(\phi \mathbf{W}) \end{aligned} \quad (11)$$

Here,  $\mu$  is the penalty factor ( $\mu$  is big enough that the algorithm converges) and  $\phi$  is the Lagrange multiplier for

constraining  $\mathbf{W}$  to be nonnegative. We consider the following situations

- (1) Considering  $\mathbf{W}$  to be fixed, we can rewrite eq 11 as a function of  $\mathbf{P}$ .

$$\min F_1(\mathbf{P}) = \|\mathbf{C}_n \mathbf{X} \mathbf{P} - \mathbf{C}_n \mathbf{X} \mathbf{W} \mathbf{P}\|_F^2 + \frac{\mu}{2} \|\mathbf{P}^T \mathbf{P} - \mathbf{I}\|_F^2 \quad (12)$$

The first and second derivatives of  $L_1$  with respect to  $\mathbf{P}$  are as follows

$$\frac{\partial F_1}{\partial \mathbf{P}} = 2(\mathbf{I} - \mathbf{W})^T \mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) \mathbf{P} + 2\mu \mathbf{P} (\mathbf{P}^T \mathbf{P} - \mathbf{I}) \quad (13)$$

$$\frac{\partial^2 F_1}{\partial \mathbf{P}^2} = 2(\mathbf{I} - \mathbf{W})^T \mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) + 2\mu (3\mathbf{P} \mathbf{P}^T - \mathbf{I}) \quad (14)$$

Define

$$\mathbf{D}_{ii} = \frac{((2(\mathbf{I} - \mathbf{W})^T \mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) + 2\mu (3\mathbf{P} \mathbf{P}^T - \mathbf{I})) \mathbf{P}^k)_i}{\mathbf{P}_i^k}$$

$$i = 1, \dots, m$$

Here,  $\mathbf{P}^k$  is the  $k$ th column of matrix  $\mathbf{P}$ . Diagonal matrix  $\mathbf{D}$  is composed of  $\mathbf{D}_{ii}$  as diagonal elements. The auxiliary function of creating  $F_1(\mathbf{P})$  with diagonal matrix  $\mathbf{D}$  is  $G(\mathbf{P}, \mathbf{P}^k) = F_1(\mathbf{P}^k) + (\mathbf{P} - \mathbf{P}^k) \nabla F_1(\mathbf{P}^k) + \frac{1}{2} (\mathbf{P} - \mathbf{P}^k)^T \mathbf{D} (\mathbf{P} - \mathbf{P}^k)$ . The update rule generated by  $\mathbf{P}^{k+1} = \arg \min_{\mathbf{P}} G(\mathbf{P}, \mathbf{P}^k)$  is as follows

$$\mathbf{P}_{ij}^{k+1} \leftarrow \mathbf{P}_{ij}^k \frac{(2\mu \mathbf{P}^k)_{ij}}{((\mathbf{I} - \mathbf{W})^T \mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) \mathbf{P}^k + \mu (3\mathbf{P} \mathbf{P}^T - \mathbf{I}) \mathbf{P}^k)_{ij}} \quad (15)$$

- (2) Considering  $\mathbf{P}$  to be fixed, we can rewrite eq 11 as a function of  $\mathbf{W}$ .

$$\begin{aligned} \min F_2(\mathbf{W}) &= \|\mathbf{C}_n \mathbf{X} \mathbf{P} - \mathbf{C}_n \mathbf{X} \mathbf{W} \mathbf{P}\|_F^2 \\ &+ \alpha \text{tr}(\mathbf{W}^T \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} \mathbf{W}), + b \|\mathbf{W}\|_{2,1} \\ &+ \frac{m}{2} (\|\mathbf{W}^T \mathbf{W} - \mathbf{I}\|_F^2 + \text{tr}(f \mathbf{W})) \end{aligned} \quad (16)$$

Here,  $\|\mathbf{W}\|_{2,1} = \mathbf{W}^T \mathbf{U} \mathbf{W}$ ,  $\mathbf{U} = \text{diag}\left(\frac{1}{2\|\mathbf{w}^i\|_2}\right)$  and  $\mathbf{w}^i$  is the  $i$ th row of matrix  $\mathbf{W}$ . With  $\mathbf{P}$  and  $\mathbf{U}$  fixed, calculate the first and second derivatives of  $F_2(\mathbf{W})$  with respect to  $\mathbf{W}$  as follows

$$\begin{aligned} \frac{\partial F_2}{\partial \mathbf{W}} &= -2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) \mathbf{P} \mathbf{P}^T + 2\alpha \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} \mathbf{W} \\ &+ 2\beta \mathbf{U} \mathbf{W} + 2\mu (\mathbf{W} (\mathbf{W}^T \mathbf{W} - \mathbf{I}) + \phi) \end{aligned} \quad (17)$$

$$\begin{aligned} \frac{\partial^2 F_2}{\partial \mathbf{W}^2} &= 2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 2\alpha \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} + 2\beta \mathbf{U} \\ &+ 2\mu (3\mathbf{W} \mathbf{W}^T - \mathbf{I}) \end{aligned} \quad (18)$$

Define the following by the second derivative.

$$\mathbf{D}_{ii}^1 = \frac{((2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 2\alpha \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} + 2\beta \mathbf{U} + 2\mu (3\mathbf{W} \mathbf{W}^T - \mathbf{I})) \mathbf{W}^k)_i}{\mathbf{W}_i^k}$$

$$i = 1, \dots, m$$

Here,  $\mathbf{w}^k$  is the  $k$ th column of matrix  $\mathbf{W}$ . Diagonal matrix  $\mathbf{D}^1$  is composed of  $\mathbf{D}_{ii}^1$  as the diagonal element. The auxiliary function of creating  $F_2(\mathbf{W})$  with diagonal matrix  $\mathbf{D}^1$  is

$$G_1(\mathbf{W}, \mathbf{W}^k) = F_2(\mathbf{W}^k) + (\mathbf{W} - \mathbf{W}^k) \nabla F_2(\mathbf{W}^k) + \frac{1}{2} (\mathbf{W} - \mathbf{W}^k)^T \mathbf{D}^1 (\mathbf{W} - \mathbf{W}^k)$$

The update rule generated by  $\mathbf{W}^{k+1} = \arg \min_p G_1(\mathbf{W}, \mathbf{W}^k)$  is as follows

$$\mathbf{W}_{ij}^{k+1} \leftarrow \mathbf{W}_{ij}^k \frac{(2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 4\mu \mathbf{W})_{ij}}{((2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 2\alpha \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} + 2\beta \mathbf{U} + 2\mu(3\mathbf{W} \mathbf{W}^T - \mathbf{I})) \mathbf{W}^k)_i} \quad (19)$$

The following FLMSR algorithm is given by the above analysis process (see Table 7).

The load matrix obtained by the FLMSR algorithm above is denoted as  $\mathbf{P}_F$ . For a newly measured sample data  $\mathbf{x}_{\text{new}} \in \mathbf{R}^t$ . Calculate the value of new sample point statistics  $t_{F,\text{new}}^2 = \mathbf{x}_{\text{new}} \mathbf{P}_F \Lambda^{-1} \mathbf{P}_F^T \mathbf{x}_{\text{new}}^T$ , where  $\Lambda = \frac{\mathbf{T}^T \mathbf{T}}{n-1} = \frac{1}{n-1} \mathbf{P}_F^T \mathbf{X}^T \mathbf{X} \mathbf{P}_F$  and statistics  $\text{spe}_{F,\text{new}} = \|\mathbf{x}_{\text{new}} (\mathbf{I} - \mathbf{P}_F \mathbf{P}_F^T)\|$ . We compare the value of statistics with the corresponding control limit to judge whether there is a fault.

### 3. CONVERGENCE OF FLMSR

The convergence of the FLMSR algorithm is discussed as follows.

**Theorem 1.** Minimization problem<sup>12</sup> is nonincreasing under the update rules

$$\mathbf{P}_{ij}^{k+1} \leftarrow \mathbf{P}_{ij}^k \frac{(2\mu \mathbf{P}^k)_{ij}}{((\mathbf{I} - \mathbf{W})^T \mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) \mathbf{P}^k + \mu(3\mathbf{P} \mathbf{P}^T - \mathbf{I}) \mathbf{P}^k)_{ij}} \quad (20)$$

**Proof.** According to the definition and the construction method of the auxiliary function in ref 34, it is concluded that the function  $G(\mathbf{P}, \mathbf{P}^k)$  defined above is the auxiliary function of function  $F_1(\mathbf{P})$ . According to ref 34, Lemma 1 finds the minimum  $\mathbf{P}^{k+1} = \arg \min_p G(\mathbf{P}, \mathbf{P}^k)$  for the auxiliary function  $G(\mathbf{P}, \mathbf{P}^k)$ , and the generated update rule is obtained as follows.

$$\mathbf{P}_{ij}^{k+1} \leftarrow \mathbf{P}_{ij}^k \frac{(2\mu \mathbf{P}^k)_{ij}}{((\mathbf{I} - \mathbf{W})^T \mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} (\mathbf{I} - \mathbf{W}) \mathbf{P}^k + \mu(3\mathbf{P} \mathbf{P}^T - \mathbf{I}) \mathbf{P}^k)_{ij}}$$

The update rule satisfies in output  $L_1(\mathbf{P}^{k+1}) \leq G(\mathbf{P}^{k+1}, \mathbf{P}^k) \leq G(\mathbf{P}^k, \mathbf{P}^k) = L_1(\mathbf{P}^k)$ .

According to equations  $L(\mathbf{P}^{k+1}, \mathbf{W}^k) = L_1(\mathbf{P}^{k+1})$  and  $L(\mathbf{P}^k, \mathbf{W}^k) = L_1(\mathbf{P}^k)$ , in output  $L(\mathbf{P}^{k+1}, \mathbf{W}^k) \leq L(\mathbf{P}^k, \mathbf{W}^k)$  can be obtained.

**Theorem 2.** Minimization problem<sup>16</sup> is nonincreasing under the update rules

$$\mathbf{W}_{ij}^{k+1} \leftarrow \mathbf{W}_{ij}^k \frac{(2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 4\mu \mathbf{W})_{ij}}{((2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 2\alpha \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} + 2\beta \mathbf{U} + 2\mu(3\mathbf{W} \mathbf{W}^T - \mathbf{I})) \mathbf{W}^k)_i} \quad (21)$$

**Proof.** Similar to the proof of theorem 1, the function  $G_1(\mathbf{W}, \mathbf{W}^k)$  is an auxiliary function of the function  $F_2(\mathbf{W})$ . Find the minimum  $\mathbf{W}^{k+1} = \arg \min_p G_1(\mathbf{W}, \mathbf{W}^k)$  for the auxiliary function  $G_1(\mathbf{W}, \mathbf{W}^k)$ , and the generated update rule is obtained as follows.

$$\mathbf{W}_{ij}^{k+1} \leftarrow \mathbf{W}_{ij}^k \frac{(2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 4\mu \mathbf{W})_{ij}}{((2\mathbf{X}^T \mathbf{C}_n^T \mathbf{C}_n \mathbf{X} \mathbf{P} \mathbf{P}^T + 2\alpha \mathbf{Y}^T (\mathbf{L}_w - \mathbf{L}_b) \mathbf{Y} + 2\beta \mathbf{U} + 2\mu(3\mathbf{W} \mathbf{W}^T - \mathbf{I})) \mathbf{W}^k)_i}$$

The update rule satisfies in output

$$L_2(\mathbf{W}^{k+1}) \leq G_1(\mathbf{W}^{k+1}, \mathbf{W}^k) \leq G(\mathbf{W}^k, \mathbf{P}^k) = L_2(\mathbf{W}^k)$$

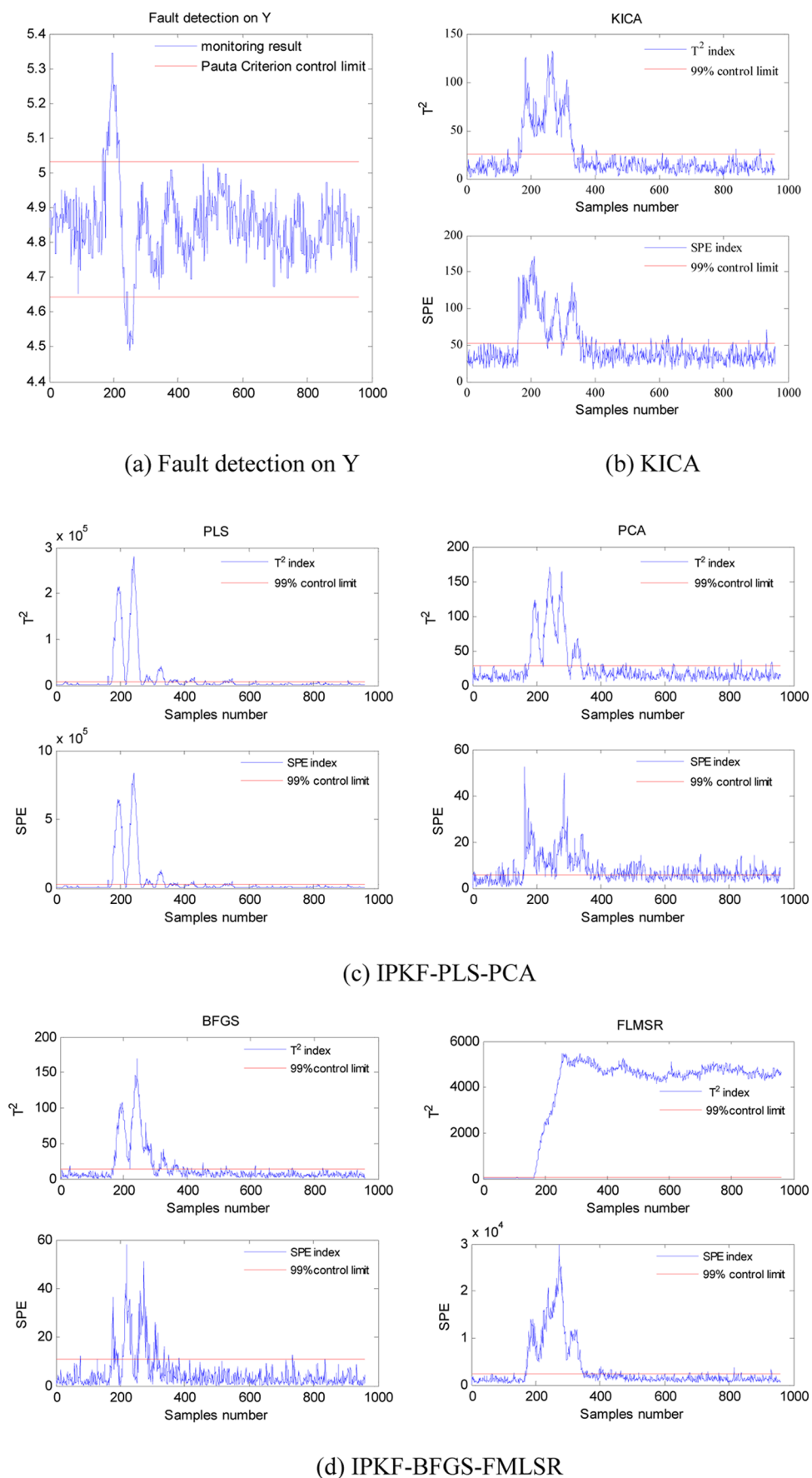
According to equations  $L(\mathbf{P}^{k+1}, \mathbf{W}^{k+1}) = L_2(\mathbf{W}^{k+1})$  and  $L(\mathbf{P}^{k+1}, \mathbf{W}^k) = L_2(\mathbf{W}^k)$ , in output  $L(\mathbf{P}^{k+1}, \mathbf{W}^{k+1}) \leq L(\mathbf{P}^{k+1}, \mathbf{W}^k)$  can be obtained.

In output  $L(\mathbf{P}^{k+1}, \mathbf{W}^{k+1}) \leq L(\mathbf{P}^{k+1}, \mathbf{W}^k) \leq L(\mathbf{P}^k, \mathbf{W}^k)$  is obtained by theorems 1 and 2, which show that the algorithm proposed in the paper is a descent iterative algorithm.

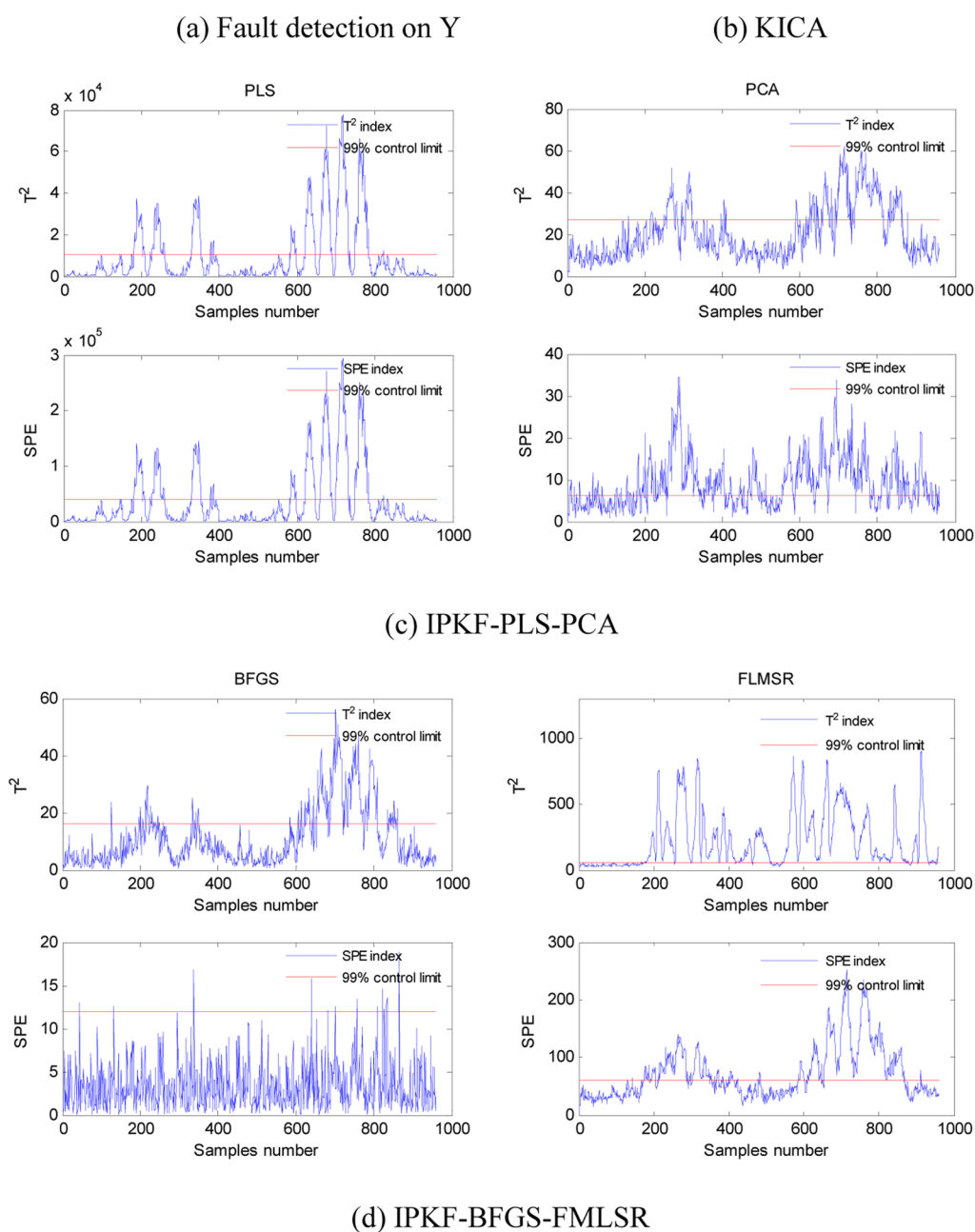
### 4. RESULTS AND DISCUSSION

The TEP was first developed by Downs and Vogel (J. J. Downs 1993) and has become a benchmark platform for process monitoring and diagnosis technique validation, which contains five major parts (i.e., a reactor, a vapor–liquid separator, a product condenser, a recycle compressor, and a product stripper). Furthermore, there are 41 measured variables, that is, XMEAS (1–41), and 12 manipulated variables, that is, XMV (1–12), in the TE process. The simulation data blocks in this paper are widely recognized in process monitoring.

By using the preprogrammed faults (faults 1–21), 21 testing blocks were generated. Fault 0 (with no faults) was generated under NOC. The testing data block for each fault contains 960 observations. Each data block starts with no faults, and the faults come across after the 160th sample. For process monitoring, 41 measurable variables XMEAS (1–41) and 11 manipulated variables XMV (1–11) are selected to construct the past data vector. Among them, the process measurements XMEAS (1–34, 36–41) and the manipulated variable XMV (1–11) together constitute the input matrix  $\mathbf{X}$ , and XMEAS (35) indicates the product quality of ingredient  $G$ , which represents the output matrix  $\mathbf{Y}$ . The following selection criteria of simulation experiment parameters are as follows: the FLMSR, PLS, and PCA. Taking the number of principal components according to the cumulative contribution rate of 80%, the number of iterations of BFGS is 80% of the number of input data columns, and KICA retains 12 independent components. The false alarm rates are maintained at the same level (99%) comparing FAR and FDR in different circumstances throughout the study.



**Figure 2.** Fault detection results of IDV (5) by three methods. (a) Fault detection on Y, (b) KICA, (c) IPKF-PLS-PCA, and (d) IPKF-BFGS-FMLSR.



**Figure 3.** Fault detection results of IDV (10) by three methods. (a) Fault detection on Y, (b) KICA, (c) IPKF-PLS-PCA, and (d) IPKF-BFGS-FMLSR.

First, the performances in fault detection of KICA, IPKF-PLS-PCA, and IPKF-BFGS-FMLSR are judged by distinguishing whether the fault occurred has an influence on the output or not. The simulation results of the faults IDV (5), IDV (10), and IDV (14) by KICA, IPKF-PLS-PCA, and IPKF-BFGS-FMLSR are shown in Figures 2–4, respectively.

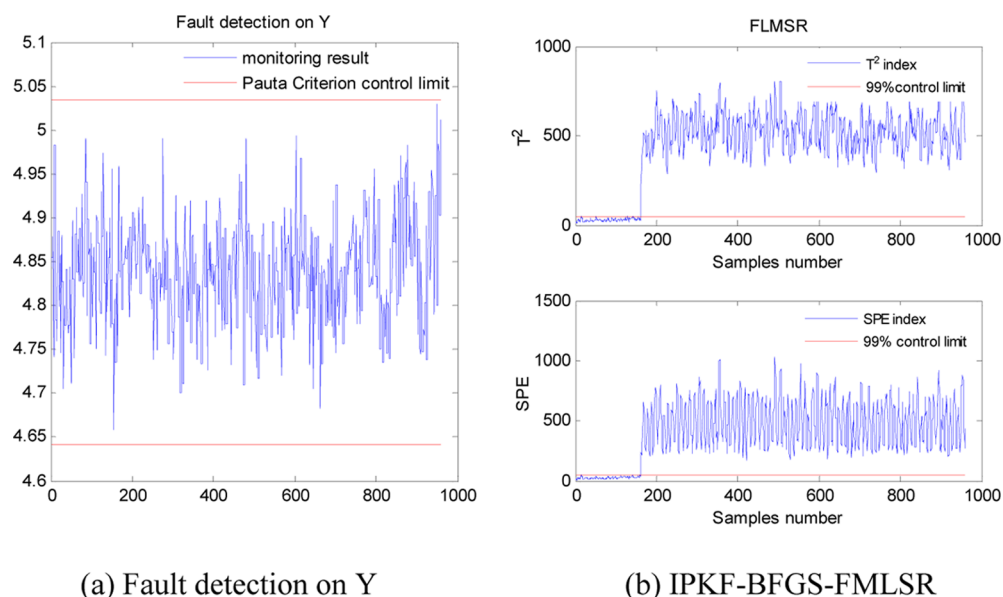
IDV (5) is a step change of condenser cooling water inlet temperature. Figure 2a shows the influence on the output caused by fault IDV (5), which takes effect from the 161st sample. It will return to a steady state after the 480th sample. It can be seen from Figure 2b,c that IDV (5) has a slight influence on the output, and the fault disappears after a period of time. The result of the BFGS block in Figure 2d indicates that IDV (5) has an influence on the output only in a very limited time. The result of the FMLSR block indicates that the

TE process is still in the abnormal operation state defined by IDV (5), even if the output is not affected. The IPKF-BFGS-FMLSR method proposed in this paper is the only method to provide correct monitoring for IDV (5).

IDV (10) occurs in the random disturbance of C feed temperature of stream 2. It can be seen from Figure 3a that IDV (10) is a fault with a slightly related output. Figure 3b–d describes the detection effects of three methods on IDV (10). It can be seen from the figure that only the FMLSR algorithm can detect faults near 161 sample points in time. Compared with other methods, the fault diagnosis rate is obviously improved. It further proves that IPKF-BFGS-FMLSR has good fault detection capability.

IDV (14) is the reactor cooling water valve sticking. Figure 4a shows that IDV (19) is an output-unrelated fault. The fault





**Figure 4.** Fault detection results of IPKF-BFGS-FMLSR for IDV (14). (a) Fault detection on Y and (b) IPKF-BFGS-FMLSR.

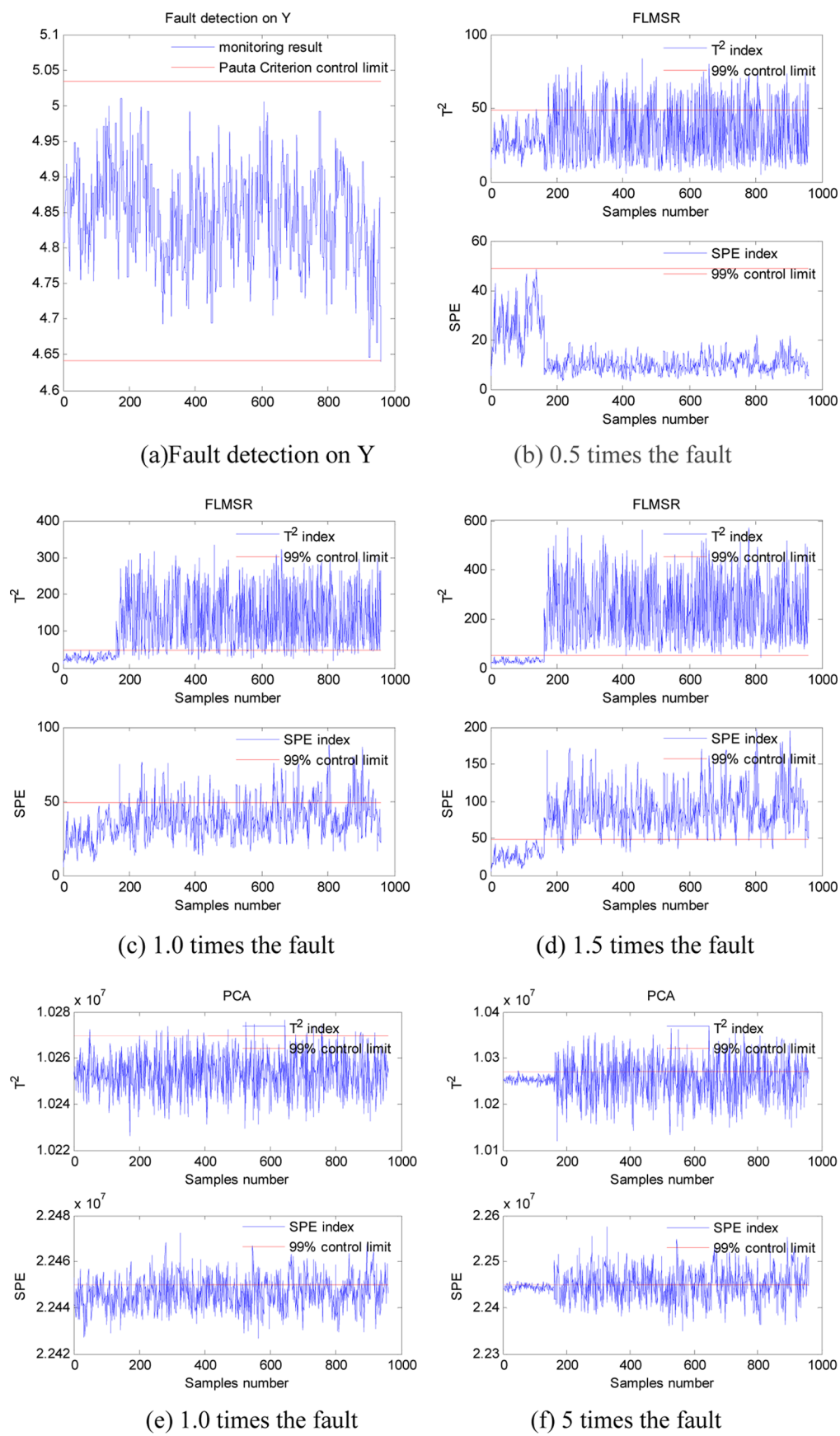
detection of IDV (14) was carried out using IPKF-BFGS-FMLSR, and it is found that the output-related block is empty. After the fault-unrelated block is removed, all variables become output-unrelated variables, and its monitoring results are shown in Figure 4b. It can be seen that there is no false report in the IPKF-BFGS-FMLSR method, and the fault diagnosis rate can be 100%. Therefore, the IPKF-BFGS-FMLSR method can track the output-unrelated faults much better, which further proves that the IPKF-BFGS-FMLSR method can identify the output-related and output-unrelated faults correctly.

IDV (19) is an unknown fault. Figure 5a shows that IDV (19) is an output-unrelated fault. In order to verify the monitoring effect of the FLMSR algorithm on the faults of different sizes, we take the average value of normal data subtracted from fault data as the fault size and construct monitoring data of different fault sizes by adjusting the fault size. Figure 5b,c,d shows the monitoring results of the original fault data, reducing the fault to half of the original fault, and enlarging the fault to 1.5 times of the original fault, respectively. It can be seen from the results that FLMSR is highly sensitive to the size of faults, and all fault data can be monitored if the faults are slightly enlarged. The fault monitoring effect is very good. Figure 5e,f shows the monitoring results of original fault data and 5 times fault data when the PCA algorithm acts on the same data block. It can be seen from the figure that although the amplifying fault has a certain effect on fault detection, it is far less sensitive to the fault size than the FLMSR algorithm.

Table 8 lists the monitoring effects of different methods. The first 21 lines reflect the fault detection rate, and line 22 reflects the average fault detection rate. Here, strictly according to whether the output exceeds the Pauta Criterion control line and whether the fault is output-related is defined. The first block in the table is the output-related fault. From the corresponding columns of BFGS and PLS algorithms in this block, it can be seen that the detection rate of the BFGS algorithm is higher than that of the PLS algorithm except IDV (10). In the whole first block, except IDV (7–8) and IDV(21), the highest monitoring rate is given by the FLMSR algorithm

proposed in this paper. The second block is the output-unrelated fault. According to the corresponding columns of BFGS and PLS algorithms in this block, it can be seen that the fault detection rate of the FLMSR algorithm proposed in this paper is higher than that of the PCA algorithm when the cumulative contribution rate is 80%. Except IDV (3) and IDV(9), the highest monitoring rate is given by the FLMSR algorithm proposed in this paper. The KICA algorithm is a very good algorithm for fault detection. The fault detection rate of IPKF-BFGS-FMLSR proposed in this paper is better than that of the KICA algorithm in most cases. In addition, obviously, the lowest false alarm rate of IPKF-BFGS-FMLSR can be observed in the last row of the table. Generally speaking, by comparing the fault detection rate and false alarm rate, IPKF-BFGS-FMLSR is found to be superior to the other two methods.

We further compare the SDR of the IPKF-BFGS-FMLSR method with that of the KICA method in Figure 6. The SDR of the IPKF-BFGS-FMLSR method is obtained by taking the maximum SDR of BFGS and FLMSR. For 21 faults, the performance of the method based on IPKF-BFGS-FMLSR is better than that based on KICA (except fault 9 and fault 12). The fault detection performance of fault 9 and fault 12 is very close, but there is still a slight difference. Taking fault 5 as an example, the SDR of the IPKF-BFGS-FMLSR method is 0.998 while that of KICA is 0.406, which shows that the SDR of the IPKF-BFGS-FMLSR method is 2.46 times higher than that of KICA. Taking fault 18 as an example, the SDR of the IPKF-BFGS-FMLSR method is 0.926 while that of KICA is 0.383, which shows that the SDR of the IPKF-BFGS-FMLSR method is also 2.4 times higher than that of KICA. The average SDR of the IPKF-BFGS-FMLSR method is 0.845 while that of KICA is 0.689, which shows that the average SDR of the IPKF-BFGS-FMLSR method is 1.2 times higher than that of KICA. This shows that FLMSR method has stronger data recognition ability. SDR results further illustrate the superior performance of the proposed method based on IPKF-BFGS-FMLSR in detecting process faults.



**Figure 5.** Fault detection results of IPKF-BFGS-FMLSR and PCA for IDV (19). (a) Fault detection on Y, (b) 0.5 times the fault, (c) 1.0 times the fault, (d) 1.5 times the fault, (e) 1.0 times the fault, and (f) 5 times the fault.

Table 7. FLMSR Algorithm

algorithm: FLMSR	
1.	Input data $X \in \mathbb{R}^{n \times t}$ , $Y \in \mathbb{R}^{2n \times t}$ parameters $\alpha$ , $\beta$ , $\sigma$ , and $\mu$ are assigned with an initial value, maximum iteration number $K$ , and iteration number counter $k = 1$ .
2.	Assign an initial random matrix between (0, 1) to the matrices $P$ and $W$ and calculate $L_w$ , $L_b$ , and $U$ .
3.	Update the $P$ matrix. $P_{ij}^{k+1} \leftarrow P_{ij}^k \frac{(2\mu P^k)_{ij}}{((1-w)^T X^T C_n^T C_n X(1-w)P^k + \mu(3PP^T - I)P^k)_{ij}}$
4.	Update the $W$ matrix.
	$W_{ij}^{k+1} \leftarrow W_{ij}^k \frac{(2X^T C_n^T C_n XPP^T + 4\mu W)_{ij}}{((2X^T C_n^T C_n XPP^T + 2\alpha Y^T(L_w - L_b)Y + 2\beta U + 2\mu(3WW^T - I)W^k)_{ij}}$
5.	Update $U = \text{diag}\left(\frac{1}{2\ w^i\ _2}\right)$ , judging that if the maximum iteration number $K$ is reached, then stop iteration, otherwise, block $k = k + 1$ and return to step 3.

Table 8. FDRs of the 21 Faults in the TE Benchmark (%)

Fault	IPKF-BFGS-FLMSR				IPKF-PLS-PCA				KICA	
	$T_B^2$	$SPE_B$	$T_F^2$	$SPE_F$	$T_{pls}^2$	$SPE_{pls}$	$T_{pca}^2$	$SPE_{pca}$	$T^2$	$SPE$
1	36.25	8.25	100	99.50	36.50	36.25	99.25	99.75	99.63	99.86
2	98.25	95.25	98.75	98.75	69.63	81.75	98.75	97.00	98.63	98.75
5	23.25	12.13	100	34.50	19.00	19.00	23.63	21.50	22.13	26.88
6	99.00	98.75	100	100	94.75	94.75	99.13	100	100	100
7	100	97.63	43.63	46.00	32.50	32.50	44.75	24.00	100	100
8	97.75	89.88	97.75	96.88	80.75	80.75	96.25	75.38	99.25	98.38
10	23.86	1.38	90.25	63.13	30.13	30.13	31.63	34.38	63.00	86.88
12	70.75	83.75	96.88	94.50	67.88	67.88	93.25	64.75	94.88	95.88
13	95.25	90.63	96.25	95.88	83.00	83.00	94.16	88.63	95.50	95.38
16	10.13	2.63	90.25	15.88	6.13	6.13	6.88	19.75	24.88	19.38
17	15.63	0.50	97.38	92.00	13.50	13.00	80.88	94.75	82.75	96.00
18	89.25	23.75	90.88	90.25	87.88	88.25	89.16	90.25	24.00	25.25
20	17.88	2.00	91.25	67.25	1.25	1.25	45.63	47.63	44.00	57.88
21	52.50	32.50	4.16	1.00	23.63	23.63	16.75	1.75	45.75	39.38
3	1.63	1.38	5.63	3.00	1.88	1.88	1.16	1.63	13.50	0.38
4	3.36	2.50	100	99.88	9.13	9.38	100	10.38	69.25	100
9	0.63	0.88	6.25	0.75	0.875	0.88	0.75	1.25	19.13	20.50
11	5.00	0.50	82.25	69.38	19.38	17.25	53.88	66.25	46.63	77.13
14	0	0	100	100	0	0	100	99.88	99.38	100
15	25.63	1.00	15.00	8.75	24.25	24.25	10.25	1.50	2.75	3.50
19	0	0	89.38	19.63	0	0	21.63	8.63	18.13	51.88
AVG	41.24	30.73	76.00	61.76	33.43	33.90	57.51	49.95	60.15	66.35
AVG-FAR	0.744	0.804	0.595	0.655	0.863	0.833	0.714	0.863	5.149	2.708

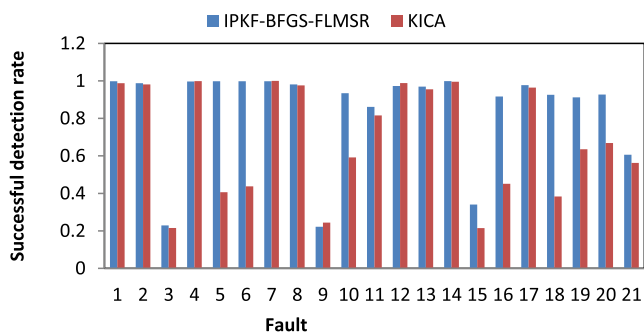


Figure 6. IPKF-BFGS-FLMSR and KICA are used for the SDR histogram of detection results of all 21 faults.

## 5. CONCLUSIONS

In this paper, a new output-related and -unrelated monitoring scheme based on modeling strategies of the IPKF algorithm, BFGS algorithm, and FLMSR algorithm is proposed. The proposed modeling program first uses IPKF to divide the input variables into fault-related blocks, output-related blocks, and

output-unrelated blocks to obtain output-related and -unrelated changes, respectively. The output-related blocks are monitored by the BFGS algorithm with better numerical stability, and the output-unrelated blocks are monitored by the FLMSR algorithm proposed in this paper. With TEP data, the IPKF-BFGS-FLMSR method, the IPKF-PLS-PCA method, and the most advanced KICA method are compared. The results show that the IPKF-BFGS-FLMSR method proposed in this paper can achieve better fault-monitoring results in output-related and -unrelated fault monitoring.

When faults occur in the industrial process, the algorithm can greatly reduce the fault range. It is not necessary to consider all variables and only need to detect output-related blocks or output-unrelated blocks to find the fault root. However, the specific fault location needs further study.

## AUTHOR INFORMATION

### Corresponding Author

Cuiping Xue – College of Science, Northeastern University, Shenyang 110819, China; [orcid.org/0000-0002-7764-5240](https://orcid.org/0000-0002-7764-5240); Phone: +86-0-18642090501; Email: [xuecuiping@mail.neu.edu.cn](mailto:xuecuiping@mail.neu.edu.cn)

## Authors

Tie Zhang – College of Science, Northeastern University, Shenyang 110819, China

Dong Xiao – College of Information Science and Engineering and Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China; [orcid.org/0000-0002-0401-6654](https://orcid.org/0000-0002-0401-6654)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.1c00506>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the State Key Laboratory of Synthetical Automation for Process Industries Fundamental Research Funds of China (no. 2013ZCX02).

## REFERENCES

- (1) Nadler, B.; Coifman, R. R. The prediction error in CLS and PLS: The importance of feature selection prior to multivariate calibration. *J. Chemom.* **2005**, *19*, 107–118.
- (2) Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **2002**, *16*, 119–128.
- (3) Zhou, D.; Li, G.; Qin, S.-J. Total projection to latent structures for process monitoring. *AIChE J.* **2010**, *56*, 168–178.
- (4) Qin, S. J.; Zheng, Y. Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures. *AIChE J.* **2013**, *59*, 496–504.
- (5) Yin, S.; Zhu, X.; Kaynak, O. Improved PLS Focused on Key-Performance-Indicator-Related Fault Diagnosis. *IRE Trans. Ind. Electron.* **2015**, *62*, 1651–1658.
- (6) Yin, S.; Wang, G.; Gao, H. Data-Driven Process Monitoring Based on Modified Orthogonal Projections to Latent Structures. *IEEE Trans. Control. Syst. Technol.* **2016**, *24*, 1480–1487.
- (7) Li, W.; Yue, H. H.; Valle-Cervantes, S.; Qin, S. J. Recursive PCA for adaptive process monitoring. *J. Process Control* **2000**, *10*, 471–486.
- (8) Li, W.; Qin, S. J. Consistent dynamic PCA based on errors-in-variables subspace identification. *J. Process Control* **2001**, *11*, 661–678.
- (9) Lee, J.-M.; Yoo, C.; Choi, S. W.; Vanrolleghem, P. A.; Lee, I.-B. Nonlinear process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **2004**, *59*, 223–234.
- (10) Chen, J.; Hou, A.; Hou, T. Y. A Prototype Knockoff Filter for Group Selection with FDR Control. *Inf. Inference* **2019**, *9*, 271–288.
- (11) Aşırım, Ö. E.; Yolalmaz, A.; Kuzuoglu, M. High-Fidelity Harmonic Generation in Optical Micro-Resonators Using BFGS Algorithm. *Micromachines* **2020**, *11*, 686.
- (12) Borhani, M. Multi-label Log-Loss function using L-BFGS for document categorization. *Eng. Appl. Artif. Intell.* **2020**, *91*, 103623.
- (13) Livieris, I. E. An advanced active set L-BFGS algorithm for training weight-constrained neural networks. *Neural. Comput. Appl.* **2020**, *32*, 6669–6684.
- (14) Meng, Y.; Shang, R.; Jiao, L.; Zhang, W.; Yuan, Y.; Yang, S. Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering. *Neurocomputing* **2018**, *290*, 87–99.
- (15) Gu, B.; Sheng, V. S. A Robust Regularization Path Algorithm for  $\nu$ -Support Vector Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 1241–1248.
- (16) Stolk, R.; Greig, A.; Hodgetts, M.; Gilby, J. An EM/E-MRF algorithm for adaptive model based tracking in extremely poor visibility. *Image Vis. Comput.* **2008**, *26*, 480–495.
- (17) Mitra, P.; Murthy, C. A.; Pal, S. K. Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 301–312.
- (18) Mao, Q.; Tsang, I. W.-H. A feature selection method for multivariate performance measures. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2051–2063.
- (19) Gu, B.; Sheng, V. S.; Wang, Z.; Ho, D.; Osman, S.; Li, S. Incremental learning for  $\nu$ -Support Vector Regression. *Neural Netw.* **2015**, *67*, 140–150.
- (20) Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. *J. Bioinf. Comput. Biol.* **2005**, *03*, 185–205.
- (21) Tian, Q.; Chen, S. Cross-heterogeneous-database age estimation through correlation representation learning. *Neurocomputing* **2017**, *238*, 286–295.
- (22) Abdulla, W. H.; Kasabov, N. Reduced feature-set based parallel CHMM speech recognition systems. *Inf. Sci.* **2003**, *156*, 21–38.
- (23) Shang, R.; Liu, C.; Meng, Y.; Jiao, L.; Stolkin, R. Nonnegative matrix factorization with rank regularization and hard constraint. *Neural Comput.* **2017**, *29*, 2553–2579.
- (24) Robnik-Šikonja, M.; Kononenko, I. Theoretical and empirical analysis of relief and relief. *Mach. Learn.* **2003**, *53*, 23–69.
- (25) Xu, Z.; King, I.; Lyu, M. R.; Jin, R. Discriminative semi-supervised feature selection via manifold regularization. *IEEE Trans. Neural Netw.* **2010**, *21*, 1033–1047.
- (26) Shi, C.; Ruan, Q.; An, G.; Zhao, R. Hessian Semi-Supervised Sparse Feature Selection Based on  $\{L_{2,1/2}\}$ -Matrix Norm. *IEEE Trans. Multimed.* **2015**, *17*, 16–28.
- (27) Li, Z.; Liu, J.; Yang, Y.; Zhou, X.; Lu, H. Clustering-guided sparse structural learning for unsupervised feature selection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *26*, 2138–2150.
- (28) Cai, D.; Zhang, C.; He, X. Unsupervised feature selection for multi-cluster data. Washington DC USA July, 2010. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery, 2010; Vol. 7; pp 333–342.
- (29) Tabakhi, S.; Parham, M.; Fardin, A. An unsupervised feature selection algorithm based on ant colony optimization. *Eng. Appl. Artif. Intell.* **2014**, *32*, 112–123.
- (30) Li, Z.; Liu, J.; Zhu, X.; Liu, T.; Lu, H. Image annotation using multi-correlation probabilistic matrix factorization. *Proceedings of the 18th International Conference on Multimedia 2010*; Association for Computing Machinery: Firenze, Italy, 2010; Vol. 10; pp 25–29.
- (31) Barber, R. F.; Candès, E.-J. Controlling the false discovery rate via knockoffs. *The Annals of Statistics. Ann. Stat.* **2015**, *43*, 2055–2085.
- (32) Barber, R.-F.; Candès, E.-J. A knockoff filter for high-dimensional selective inference. *Ann. Stat.* **2019**, *47*, 2504–2537.
- (33) Reid, S.; Tibshirani, R. Sparse regression and marginal testing using cluster prototypes. *Biostatistics* **2016**, *27*, 364–376.
- (34) Lee, D.-D.; Seung, H.-S. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems 13*; Papers from Neural Information Processing Systems (NIPS) 2000: Denver, CO, USA, 2000; Vol. 1; pp 535–541.