processes

MDPI

Article

Distributed Robust Dictionary Pair Learning and Its Application to Aluminum Electrolysis Industrial Process

Jingkun Wang [1,2], Xiaofang Chen [3], Ziqing Deng [3,*], Hongliang Zhang [1,2] and Jing Zeng [3]

[1] School of Metallurgy and Environment, Central South University, Changsha 410083, China
[2] National Engineering Research Center for Low-Carbon Nonferrous Metallurgy, Changsha 410083, China
[3] School of Automation, Central South University, Changsha 410083, China
* Correspondence: ziqingddeng@163.com; Tel.: +86-731-88830700

Abstract: In modern industrial systems, high-dimensional process data provide rich information for process monitoring. To make full use of local information of industrial process, a distributed robust dictionary pair learning (DRDPL) is proposed for refined process monitoring. Firstly, the global system is divided into several sub-blocks based on the reliable prior knowledge of industrial processes, which achieves dimensionality reduction and reduces process complexity. Secondly, a robust dictionary pair learning (RDPL) method is developed to build a local monitoring model for each sub-block. The sparse constraint with $l_{2,1}$ norm is added to the analytical dictionary, and a low rank constraint is applied to the synthetical dictionary, so as to obtain robust dictionary pairs. Then, Bayesian inference method is introduced to fuse local monitoring information to global anomaly detection, and the block contribution index and variable contribution index are used to realize anomaly isolation. Finally, the effectiveness of the proposed method is verified by a numerical simulation experiment and Tennessee Eastman benchmark tests, and the proposed method is then successfully applied to a real-world aluminum electrolysis process.

Keywords: high-dimension; distributed robust dictionary pair learning; process monitoring; aluminum electrolysis

Citation: Wang, J.; Chen, X.; Deng, Z.; Zhang, H.; Zeng, J. Distributed Robust Dictionary Pair Learning and Its Application to Aluminum Electrolysis Industrial Process. *Processes* **2022**, *10*, 1850. https://doi.org/10.3390/pr10091850

Academic Editor: Blaž Likozar

Received: 23 August 2022
Accepted: 9 September 2022
Published: 14 September 2022

1. Introduction

With the continuous development of information technology, the degree of automation and integration of industrial processes have been continuously improved, and the structure of industrial systems has become more complex. There are more and more factors affecting the stable operation of industrial production, which makes anomaly detection and isolation of industrial processes face challenges [1,2]. At present, industrial process monitoring technology has attracted great attention from academia and industry. The excellent properties of aluminum make it widely used in practical fields, including construction, electrical, packaging, medicine, transportation and so on. Electrolytic cell is the core equipment of aluminum production industry. Moreover, aluminum electrolysis process mainly depends on biggish and large pre-baked anode electrolytic cell. For the large-scale electrolytic cell, once the abnormal cell condition occurs, the efficient and stable operation of the electrolytic cell may be destroyed, causing huge economic losses [3]. Therefore, effective process monitoring technology is of profound significance to ensure the green and stable operation of electrolytic cells and reduce the production cost of enterprises. Existing aluminum electrolysis process monitoring methods can be divided into three categories: mechanistic model-based methods [4,5], knowledge-based methods [6,7], and data-driven methods [8–10]. The effectiveness of the mechanism model-based process monitoring method for aluminum electrolysis depends on the measurements, like cell resistance, cell voltage, and aluminum level. However, some process parameters such as electrolyte temperature and molecular ratio cannot be measured directly in actual aluminum electrolytic production. Meanwhile, individual differences in empirical knowledge and lack of systematic theoretical basis lead

to the failure to ensure the effectiveness of knowledge-based process monitoring methods in aluminum electrolysis. In addition, due to the reduction on the cost of sensors, the rapid development of technology and the application of advanced computer technology, modern industrial systems use a large number of sensors to obtain rich process information, and data-driven process monitoring methods are widely studied and applied [11,12].

In recent years, dictionary learning has been proposed as an effective statistical machine learning method. Compared with traditional data-driven methods, dictionary learning methods have good generalization ability, which are successfully applied in many fields, such as pattern recognition, image processing, and computer vision [13]. Dictionary learning seeks the linear combination of atoms to reconstruct the original data. And the learned dictionaries are over-complete and not restricted by orthogonal, which makes dictionary atoms adapt to the training data more flexibly to ensure the high precision of learning methods. According to different sparse coding methods, traditional dictionary learning methods include analytical dictionary learning [14,15] and synthetical dictionary learning [16,17]. Analytical dictionary learning can directly build the required dictionary from the precise and fast transformation base, and its predefined encoding method makes the computational complexity low, but it is relatively limited in monitoring modeling ability. Synthetical dictionary learning requires sparse reconstruction by l_0 or l_1 norm, which leads to much higher computational complexity. However, synthetical dictionary learning is developing in industrial process monitoring with the effective ability of local modeling. Meanwhile, to combine the advantages of analytical dictionary learning and synthetical dictionary learning, Gu et al. [18] combined analytical dictionaries and synthetical dictionaries into a learning framework and proposed dictionary pair learning (DPL) method. Zhang et al. [19] integrated coefficient learning and salient feature extraction into a unified model and proposed a self-expressed local adaptive potential dictionary pair learning method. Sun et al. [20] proposed a structured robust adaptive dictionary pair learning framework for discriminative sparse representation learning, realizing the strong representation ability of available samples.

However, small anomalies of real industrial processes are often hidden in high-dimensional data. Commonly used dimensionality reduction methods include principal component analysis (PCA) [21], partial least squares (PLS) [22], local retained projection (LPP) [23], and other multivariate statistical monitoring methods. These methods project the highly correlated high-dimensional process data into the low-dimensional subspace by selecting representative principal elements, but this way destroys the global structure of the original data matrix. In addition, traditional process monitoring methods often build global monitoring model and ignore the local behavior of industrial processes. In order to carry out regional and refined process monitoring, distributed process monitoring technologies have emerged in the era of big data [24]. Distributed process monitoring technologies combine the prior knowledge of industrial process to divide the high-dimensional data into blocks, and build monitoring sub-models for each block, then fuse the monitoring results of sub-blocks to the global monitoring results. For example, Zhu et al. [25] decomposed large-scale processes into distributed blocks with prior process knowledge, and proposed a distributed parallel data processing strategy based on MapReduce framework. Xu et al. [26] proposed a distributed principal component analysis method for angle-relevant variable selection, and realized whole plant process monitoring by the reduction of process variables and the extraction of potential features. Huang et al. [27] proposed a distributed dictionary learning (DDL) method for fault detection and fault isolation to achieve efficient industrial process monitoring. Thus, existing process monitoring methods for high-dimensional data can be divided into two categories, including projected methods and distributed learning methods. The schematic diagram of the two methods is shown in Figure 1. In fact, industrial process data are always contaminated by noise or outliers, which adversely affects process monitoring efficiency. Therefore, robust process monitoring methods are particularly important.

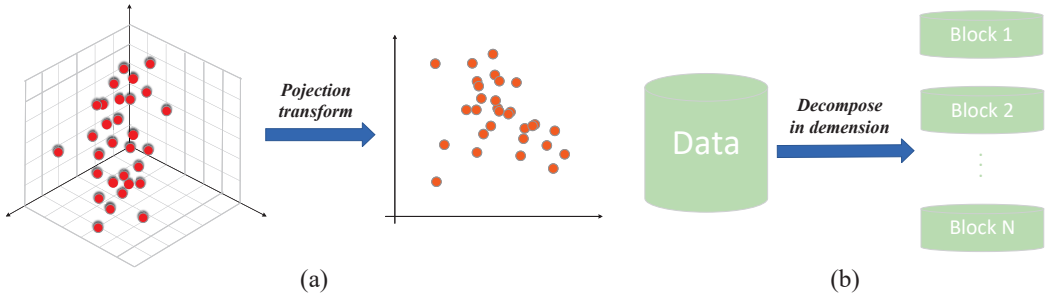

Figure 1. Schematic diagram of the two process monitoring methods for high-dimensional data. (**a**) Projected methods; (**b**) Distributed learning methods.

To solve the problem of noise and outliers in high-dimensional industrial process data, a process monitoring method based on distributed robust dictionary pair learning (DRDPL) is proposed in this article. The main contributions of this article are described as follows. The global system is divided into several sub-blocks with the reliable prior knowledge of industrial process, which achieves dimensionality reduction and reduces process complexity. A robust dictionary pair learning (RDPL) method is proposed to build local process monitoring models for each sub-block. A sparse metric based on $l_{2,1}$ norm is used to encode the reconstruction errors so as to avoid the costly computation of l_0 and l_1 norm. To reduce the interferences of noise and outliers, a sparse constraint with $l_{2,1}$ norm is added to the analytical dictionary, and a low rank constraint is applied to the synthetical dictionary, providing a robust dictionary pair. Then, Bayesian inference method is introduced to fuse local monitoring information for global anomaly detection, and the block contribution index and variable contribution index are used to realize anomaly isolation, obtaining the interpretable location results of anomaly sources.

The remainder of this article is organized as follows. Section 2 describe the proposed method in detail. Section 3 presents a numerical simulation experiment, Tennessee Eastman benchmark tests, and a real-world aluminum electrolysis process monitoring. Finally, the conclusion is given in Section 4.

2. Methodology

In this section, we will introduce the process monitoring method based on distributed robust dictionary pair learning (DRDPL) in detail. The proposed method is mainly divided into three stages, including distributed robust dictionary pair learning (RDPL), anomaly detection, anomaly isolation. In the distributed robust dictionary pair learning stage, training samples are divided in dimensions with the prior process knowledge. RDPL sub-models are built for each sub-block, learning robust synthetical dictionary, robust analytical dictionary, and control threshold. In the stage of anomaly detection, testing samples are divided into blocks with the division of training samples, then the local monitoring information obtained by the learned synthetical dictionary and analytical dictionary is fused to the global one based on Bayesian inference, so as to detect whether the testing sample is abnormal. In the anomaly isolation stage, the contribution plot method is used to define the block contribution index and variable contribution index to locate abnormal sources, where the block contribution index is used to achieve block-level anomaly isolation, and the accurate anomaly location of variables in each block is realized by the variable contribution index. The schematic diagram of the proposed process monitoring method is shown in Figure 2.

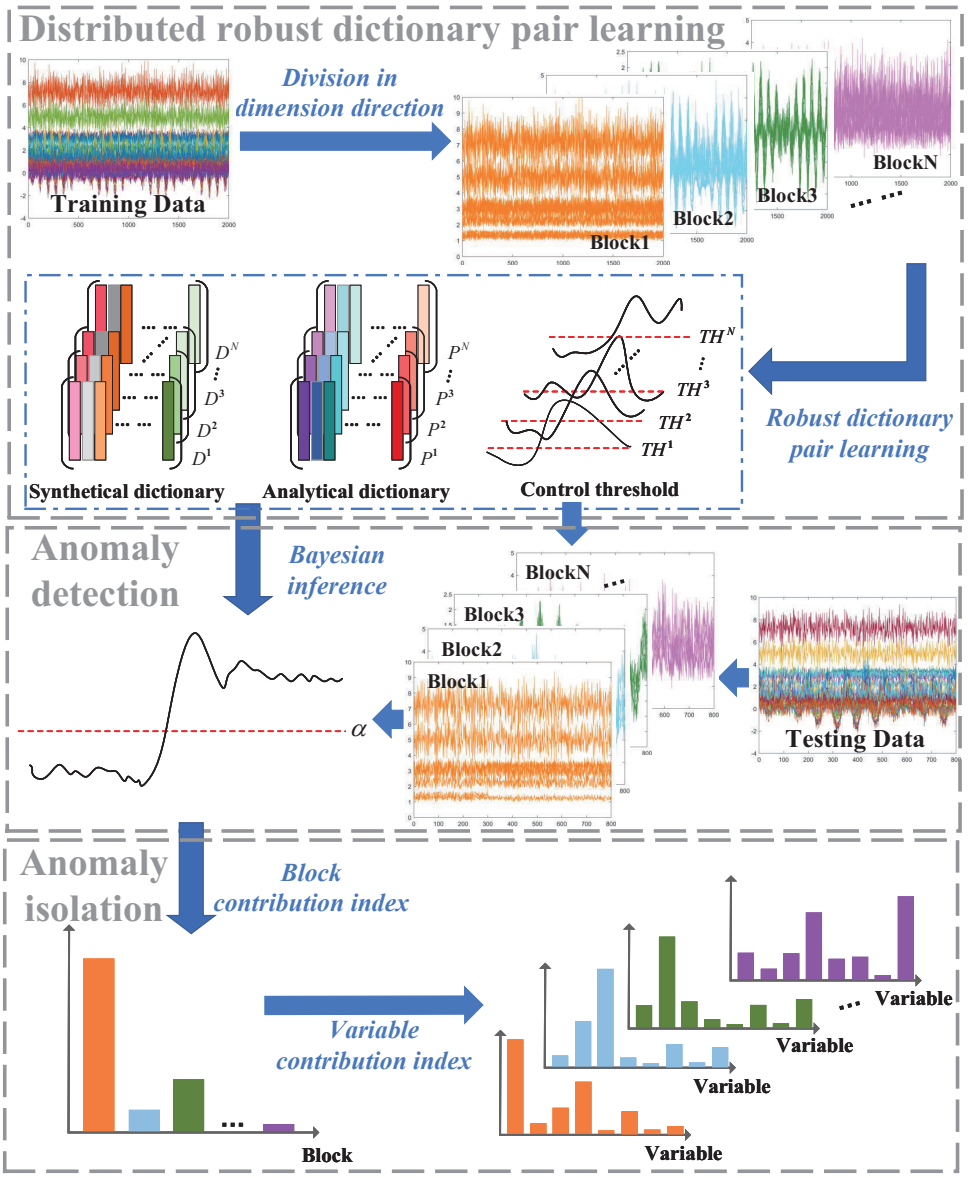

Figure 2. Schematic diagram of the process monitoring method based on distributed robust dictionary pair learning.

2.1. Distributed Robust Dictionary Pair Learning

To solve the deficiency of traditional dictionary learning, dictionary pair learning combines synthetical dictionary and analytical dictionary to reduce the computational burden of l_0 or l_1 norm constraint and enhance the reconstruction ability of dictionary learning [18]. The structure diagram of dictionary pair learning is shown in Figure 3, and its general model is formulated as follows:

$$\{D^*, P^*\} = \arg\min_{D,P} \left\{ \|X - DPX\|_F^2 + \psi(D, P, X, Y) \right\} \quad (1)$$

where $X \in R^{p \times n}$ is p-dimensional data matrix, $D \in R^{p \times m}$ represents the synthetical dictionary, $P \in R^{m \times p}$ is the analytical dictionary, m is the number of dictionary atoms. $\|X - DPX\|_F^2$ denotes the reconstruction error term of dictionary pair learning, $\psi(D, P, X, Y)$ are some discriminative functions, and Y stands for the label matrix of X. In the dictionary pair learning model, the representation coefficients A can be obtained by linear projection instead of nonlinear sparse coding with l_0 or l_1 norm. That is, we can learn an analytical dictionary P, such that A can be analytically obtained as $A = PX \in R^{m \times n}$. Based on this,

dictionary pair learning method learns such an analytical dictionary P together with the synthetical dictionary D, then the data matrix X can be reconstructed by D, P, and X, i.e., $X \approx DPX$, where D is used to reconstruct X, and P is applied to analytically code X.

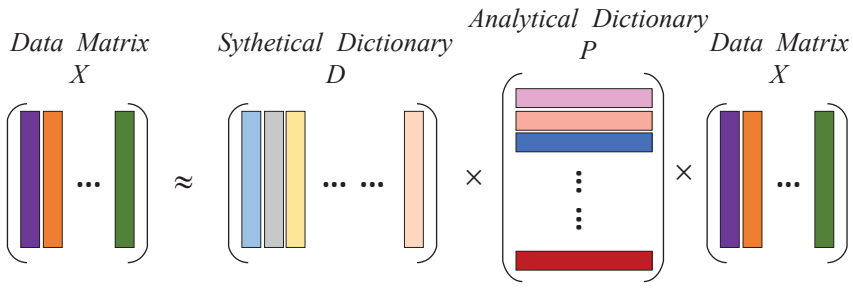

Figure 3. Structure diagram of dictionary pair learning.

Dictionary pair learning described above has been improved and used in industrial process monitoring. However, the process data collected in practical industrial systems often contain noise and outliers, which bring difficulties to process monitoring. Figure 4 shows the impact of noise and outliers on process monitoring results, where α is the control threshold, a value lower than α is considered normal, while a value higher than α is detected as an anomaly. From Figure 4a, we can observe that normal samples and abnormal samples can be correctly detected without the interferences of noise and outlier. In Figure 4b, we can see that when process data contains noise and outliers, false positives will appear in abnormal detection results of process monitoring.

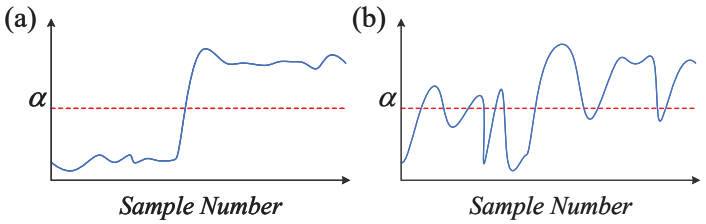

Figure 4. Illustration of the effects of noise and outliers in process monitoring. (**a**) is the training data without noise and outliers; (**b**) is the training data containing noise and outliers.

Therefore, to address the problem of process monitoring performance degradation caused by outliers and noise, we propose a robust dictionary pair learning method for industrial process monitoring. In addition, the proposed process monitoring method based on distributed robust dictionary pairs is developed for high-dimensional process data. The prior process knowledge is used to divide the training samples into blocks in the dimension direction, i.e., $X = [X_1, \cdots, X_K, \cdots, X_N]^T$. The RDPL model of the Kth block in training samples $X_K \in R^{d_K \times n}$ is denoted as follows:

$$\arg\min_{D_K, P_K} \left\{ \left\| X_K^T - X_K^T P_K D_K^T \right\|_{2,1} + \alpha \|P_K\|_{2,1} + \beta rank(D_K) \right\}, \text{s.t.} P_K^T X_K \geq 0 \tag{2}$$

where $D_K \in R^{d_K \times m}$ represents the synthetical dictionary and $P_K \in R^{d_K \times m}$ represents the analytical dictionary for the Kth block, respectively. m denotes the number of dictionary atoms. The first term is the reconstruction function of data, the second term is the sparse regularization of analytical dictionary, and the third term is the low-rank constraint of synthetical dictionary. α and β are the positive parameters used to balance the terms. Besides, the constraint $P_K^T X_K \geq 0$ is imposed to ensure that the coding coefficient $P_K^T X_K$ is non-negative.

By introducing analytical coding matrix A_K, the non-convex problem of Equation (2) is relaxed and transformed into the optimization function as follows:

$$\arg \min_{D_K, P_K, A_K} \left\{ \left\| X_K^T - A_K^T D_K^T \right\|_{2,1} + \lambda \left\| A_K^T - X_K^T P_K \right\|_{2,1} + \alpha \|P_K\|_{2,1} + \beta rank(D_K) \right\}, \text{s.t.} A_K \geq 0 \quad (3)$$

where $A_K \approx P_K^T X_K$, and λ is a scalar constant. The optimization of the objective function in Equation (3) is conducted in the following steps.

Numbered lists can be added as follows:

(1) Fix D_K and P_K, update A_K

Firstly, we fix the synthetical dictionary D_K and the analytical dictionary P_K, the problem with respect to the analytical coding matrix A_K can be reformulated as follows:

$$\arg \min_{A_K} \left\{ \left\| X_K^T - A_K^T D_K^T \right\|_{2,1} + \lambda \left\| A_K^T - X_K^T P_K \right\|_{2,1} \right\}, \text{s.t.} A_K \geq 0 \quad (4)$$

Based on the definition of $l_{2,1}$ norm [28], we have $\left\| X_K^T - A_K^T D_K^T \right\|_{2,1} = 2tr[(X_K - D_K A_K) \cdot U_K (X_K^T - A_K^T D_K^T)]$, where U_K is a diagonal matrix with the (i,i)th diagonal entries $U_K^{ii} = 1 / \left[2 \left\| \left(X_K^T - A_K^T D_K^T \right)^i \right\|_2 \right]$, $\left(X_K^T - A_K^T D_K^T \right)^i$ is the ith row vector of $X_K^T - A_K^T D_K^T$. In fact, since $\left\| \left(X_K^T - A_K^T D_K^T \right)^i \right\|_2$ may be equal to 0, we approximate $2 \left\| \left(X_K^T - A_K^T D_K^T \right)^i \right\|_2 + \tau$ instead. τ is a small value to avoid singular values and to make the inversion more stable. Similarly, $\left\| A_K^T - X_K^T P_K \right\|_{2,1} = 2tr\left[(A_K - P_K^T X_K) V_K (A_K^T - X_K^T P_K) \right]$, where V_K is a diagonal matrix with the (i,i)th diagonal entries $V_K^{ii} = 1 / \left[2 \left\| \left(A_K^T - X_K^T P_K \right)^i \right\|_2 \right]$, $\left(A_K^T - X_K^T P_K \right)^i$ is the ith row vector of $A_K^T - X_K^T P_K$. We use $2 \left\| \left(A_K^T - X_K^T P_K \right)^i \right\|_2 + \tau$ to approximate $2 \left\| \left(A_K^T - X_K^T P_K \right)^i \right\|_2$. Then, the problem with respect to A_K can be reformulated as follows:

$$\arg \min_{A_K} \left\{ \begin{array}{l} 2tr\left[(X_K - D_K A_K) U_K (X_K^T - A_K^T D_K^T) \right] \\ +2\lambda tr\left[(A_K - P_K^T X_K) V_K (A_K^T - X_K^T P_K) \right] \end{array} \right\}, \text{s.t.} A_K \geq 0 \quad (5)$$

Let $\psi_{K,rc}$ be the Lagrange multiplier for $A_{K,rc} \geq 0$ and $\Psi = [\psi_{K,rc}]$ [20], the Lagrange function ζ can be deduced as follows:

$$\begin{array}{c} \zeta = 2tr\left[(X_K - D_K A_K) U_K (X_K^T - A_K^T D_K^T) \right] \\ +2\lambda tr\left[(A_K - P_K^T X_K) V_K (A_K^T - X_K^T P_K) \right] + tr(\Psi A_K^T) \end{array} \quad (6)$$

The partial derivatives of ζ with respect to A_K in Equation (6) are computed as follows:

$$\frac{\partial \zeta}{\partial A_K} = -4 D_K^T X_K U_K + 4 D_K^T D_K A_K U_K + 4\lambda A_K V_K - 4\lambda P_K^T X_K V_K + \Psi \quad (7)$$

By the definition of KTT condition [29], the equation with respect to $A_{K,rc}$ is obtained as follows:

$$\begin{array}{c} -\left(4 D_K^T X_K U_K \right)_{rc} A_{K,rc} + 4 \left(D_K^T D_K A_K U_K \right)_{rc} A_{K,rc} + 4\lambda (A_K V_K)_{rc} A_{K,rc} \\ -4\lambda \left(P_K^T X_K V_K \right)_{rc} A_{K,rc} = 0 \end{array} \quad (8)$$

Thus, we can obtain that the elements in rth row and cth column of A_K are updated as follows:

$$A_{K,rc}^{t+1} \leftarrow A_{K,rc}^t \frac{\left(D_K^{tT} X_K U_K^t + \lambda P_K^{tT} X_K V_K^t \right)_{rc}}{\left(D_K^{tT} D_K^t A_K^t U_K^t + \lambda A_K^t V_K^t \right)_{rc}} \quad (9)$$

(2) Fix A_K and D_K, update P_K

Secondly, after the analytical coding matrix A_K is updated, we can update the analytical dictionary P_K. By removing the terms that irrelevant to P_K, the problem in Equation (3) is reformulated as follows:

$$\arg\min_{P_K}\left\{\lambda\left\|A_K^T - X_K^T P_K\right\|_{2,1} + \alpha\|P_K\|_{2,1}\right\} \tag{10}$$

Similarly, we have $\|P_K\|_{2,1} = 2tr\left(P_K^T M_K P_K\right)$, where M_K is a diagonal matrix with the (i,i)th diagonal entries $M_K^{ii} = 1/\left[2\left\|P_K^i\right\|_2\right]$, P_K^i is the ith row vector of P_K. We use $2\left\|P_K^i\right\|_2 + \tau$ to approximate $2\left\|P_K^i\right\|_2$. Then, the problem with respect to P_K can be converted as follows:

$$\arg\min_{P_K}\left\{2\lambda tr\left[\left(A_K - P_K^T X_K\right)V_K\left(A_K^T - X_K^T P_K\right)\right] + 2\alpha tr\left(P_K^T M_K P_K\right)\right\} \tag{11}$$

Let the partial derivative of Equation (11) with respect to P_K be 0, and we can obtain the closed-form solution of P_K as follows:

$$P_K^{t+1} = \left(\lambda X_K V_K^t X_K^T + \alpha M_K^t + \tau I\right)^{-1} \cdot \left(\lambda X_K V_K^t A_K^{(t+1)T}\right) \tag{12}$$

(3) Fix A_K and P_K, update D_K

Finally, after the analytical dictionary P_K is calculated, we can update the synthetical dictionary D_K. By removing the terms that irrelevant to D_K, the problem with respect to D_K is expressed as follows:

$$\arg\min_{D_K}\left\{\left\|X_K^T - A_K^T D_K^T\right\|_{2,1} + \beta rank(D_K)\right\} \tag{13}$$

Obviously, the optimization problem in Equation (13) is an NP-hard problem. Therefore, we use the low-rank function with the nuclear norm constraint to relax the optimization problem as follows [30]:

$$\arg\min_{D_K}\left\{\left\|X_K^T - A_K^T D_K^T\right\|_{2,1} + \beta\|D_K\|_*\right\} \tag{14}$$

where $\|D_K\|_*$ is the nuclear of D_K. To reduce the computational complexity, we use $\|R_K\|_F^2 + \|S_K\|_F^2$ to replace $\|D_K\|_*$ [31], where $R_K \in R^{d_K \times d_K}$ and $S_K \in R^{d_K \times m}$. Thus, the optimization problem of Equation (14) can be converted as follows:

$$\arg\min_{D_K}\left\{\left\|X_K^T - A_K^T D_K^T\right\|_{2,1} + \frac{\beta}{2}\left(\|R_K\|_F^2 + \|S_K\|_F^2\right)\right\}, \text{s.t.} D_K = R_K S_K \tag{15}$$

We use the inexact ALM algorithm [30] to solve the optimization problem in Equation (15), and the augmented Lagrange function is formulated as follows:

$$L = \left\|X_K^T - A_K^T D_K^T\right\|_{2,1} + \frac{\beta}{2}\left(\|R_K\|_F^2 + \|S_K\|_F^2\right) + \frac{\gamma}{2}\|D_K - R_K S_K\|_F^2 + \langle\mu, D_K - R_K S_K\rangle \tag{16}$$

where $\gamma > 0$ is a penalty parameter, and μ is a Lagrange multiplier.

To solve the optimization problem in Equation (16), we minimize the augmented Lagrange function by iterative updating as follows:

(1) Fix R_K and S_K, update D_K

By removing the terms of Equation (16) that irrelevant to D_K, the optimization problem with respect to D_K can be reformulated as follows:

$$\arg\min_{D_K}\left\|X_K^T - A_K^T D_K^T\right\|_{2,1} + \frac{\gamma}{2}\|D_K - R_K S_K\|_F^2 + \langle\mu, D_K - R_K S_K\rangle \tag{17}$$

Let the partial derivative of Equation (17) with respect to D_K be 0, and we can update the synthetical dictionary D_K as follows:

$$D_K^* = \left(4X_K U_K A_K^T + R_K S_K - \mu I\right) \cdot \left(4A_K U_K A_K^T + \gamma I\right)^{-1} \tag{18}$$

(2) Fix D_K and S_K, update R_K

By removing the terms of Equation (16) that irrelevant to R_K, the optimization problem with respect to R_K can be reformulated as follows:

$$\arg\min_{R_K} \frac{\beta}{2}\|R_K\|_F^2 + \frac{\gamma}{2}\|D_K - R_K S_K\|_F^2 + \langle \mu, D_K - R_K S_K \rangle \tag{19}$$

Let the partial derivative of Equation (19) with respect to R_K be 0, and we can update the variable matrix R_K as follows:

$$R_K^* = \left(\gamma D_K S_K^T + \mu S_K^T\right) \cdot \left(\beta I + \gamma S_K S_K^T\right)^{-1} \tag{20}$$

(3) Fix D_K and R_K, update S_K

By removing the terms of Equation (16) that irrelevant to S_K, the optimization problem with respect to S_K can be reformulated as follows:

$$\arg\min_{S_K} \frac{\beta}{2}\|S_K\|_F^2 + \frac{\gamma}{2}\|D_K - R_K S_K\|_F^2 + \langle \mu, D_K - R_K S_K \rangle \tag{21}$$

Let the partial derivative of Equation (21) with respect to S_K be 0, and we can update the variable matrix S_K as follows:

$$S_K^* = \left(\beta I + \gamma R_K^T R_K\right)^{-1} \cdot \left(\gamma R_K^T D_K + \mu R_K^T\right) \tag{22}$$

Thus, the iterative updating process of the synthetical dictionary D_K is summarized as follows:

$$\begin{cases} D_K^{(r+1)} = \left(4X_K U_K A_K^T + R_K^r S_K^r - \mu I\right) \cdot \left(4A_K U_K A_K^T + \gamma I\right)^{-1} \\ R_K^{(r+1)} = \left(\gamma D_K^{(r+1)} S_K^{rT} + \mu S_K^{rT}\right) \cdot \left(\beta I + \gamma S_K^r S_K^{rT}\right)^{-1} \\ S_K^{(r+1)} = \left(\beta I + \gamma R_K^{(r+1)T} R_K^{(r+1)}\right)^{-1} \cdot \left(\gamma R_K^{(r+1)T} D_K^{(r+1)} + \mu R_K^{(r+1)T}\right) \end{cases} \tag{23}$$

To fully introduce the proposed method, Algorithm 1 describes the optimization of RDPL, which stops optimizing each variable when the algorithm reaches the maximum iteration T.

Algorithm 1 Robust Dictionary Pair Learning

1: **Input:** The training samples X_K of the Kth sub-block, the parameters α, β, λ, γ, μ and τ.
2: **Step 1:** Initialize the synthetical dictionary $D_K^{(0)}$ and the analytical dictionary $P_K^{(0)}$ as random matrixes with unit Frobenius norm, set $t = 0$.
3: **Step 2:** Repeat until $t > T - 1$;
4: **Step 2.1:** Fix the analytical dictionary P_K and the synthetical dictionary D_K, update the analytical coding matrix $A_K^{(t+1)}$ by Equation (9);
5: **Step 2.2:** Fix the synthetical dictionary D_K and the analytical coding matrix A_K, update the analytical dictionary $P_K^{(t+1)}$ by Equation (12);
6: **Step 2.3:** Fix analytical coding matrix A_K and the analytical dictionary P_K, update the synthetical dictionary $D_K^{(t+1)}$ by Equation (23);
7: **Step 2.4:** Set $t = t + 1$.
8: **Output:** The analytical dictionary P_K^* and the synthetical dictionary D_K^* of the Kth sub-block.

By building RDPL model through Algorithm 1, we can calculate the reconstruction error of training samples in the Kth sub-block as follows:

$$E_K = \left\| X_K - D_K^* P_K^{*T} X_K \right\|_F^2 \tag{24}$$

Then, the control threshold C_K of the Kth sub-block can be obtained by the kernel density estimation (KDE) method [32], and the univariate kernel density estimation is conducted as follow:

$$f_H(x) = \frac{1}{MH} \sum_{i=1}^{M} K\left(\frac{x - E_K^i}{H} \right) \tag{25}$$

where x represents the data point under consideration, M is the number of training samples, H represents the bandwidth, E_K^i is the reconstruction error of the th sample in the Kth sub-block, and $K(\cdot)$ is the uniform kernel function.

2.2. Bayesian Inference Based Anomaly Detection

Testing samples Y_{new} are divided into $Y_{new} = [y_1, \cdots, y_K, \cdots, y_N]$ by the means of division in training samples, and the reconstruction error of the Kth sub-block in testing samples is calculated according to the corresponding RDPL model as follows:

$$E_{y_K} = \left\| y_K - D_K^* P_K^{*T} y_K \right\|_F^2 \tag{26}$$

To fuse the local monitoring statistics of sub-blocks to the global monitoring information in industrial processes, the Bayesian inference method [33] is introduced to convert the reconstruction error E_{y_K} of sub-block y_K in testing samples into normal possibility $P_K(y_K|N)$ and anomaly possibility $P_K(y_K|A)$, which are expressed as follows:

$$P_K(y_K|N) = e^{-\frac{E_{y_K}}{C_K}} \tag{27}$$

$$P_K(y_K|A) = e^{-\frac{C_K}{E_{y_K}}} \tag{28}$$

The conditional probability of normal sub-blocks and the conditional probability of abnormal sub-blocks are defined as $P_K(N)$ and $P_K(A)$ by significance level α, respectively, i.e., $P_K(N) = 1 - \alpha$ and $P_K(A) = \alpha$. And the posterior probability of abnormal sub-blocks based on Bayesian inference method is calculated as follows:

$$P_K(A|y_K) = \frac{P_K(y_K|A) P_K(A)}{P_K(y_K|A) P_K(A) + P_K(y_K|N) P_K(N)} \tag{29}$$

Then, the global anomaly index (GAI) is defined to fuse local statistical information to global state, which is expressed as

$$GAI = \sum_{K=1}^{N} \frac{P_K(y_K|A) p_K(A|y_K)}{\sum_{K=1}^{N} P_K(y_K|A)} \tag{30}$$

where N is the number of sub-blocks in testing samples. If a new testing sample y_{new} satisfies $GAI_{new} < \alpha$, it regards as normal, otherwise it regards as anomaly.

2.3. Contribution Index Based Anomaly Isolation

For the detected abnormal samples, we need to further locate abnormal sources. The location of the anomaly is found by the method of locating the abnormal block based

on counting time [27], that is, when the posteriori probability of the abnormal block exceeds the significance level, the block anomaly flag (BAF) is set to 1, and BAF is defined as

$$BAF_K^h = \begin{cases} 1, & P_K\left(A\middle|y_K^h\right) \geq \alpha \\ 0, & \text{Otherwise} \end{cases} \tag{31}$$

where y_K^h represents the hth abnormal sample of the Kth block. To ensure the reliability of abnormal block isolation, block anomaly index (BAI) and block contribution index (BCI) are defined as follows:

$$BAI_K = \sum_{h=1}^{H} BAF_K^h \tag{32}$$

$$BCI_K = \frac{BAI_K}{\sum_{i=1}^{N} BAI_K} \tag{33}$$

where H is the number of abnormal samples.

In industrial process monitoring, contribution plot method [34] has become a common method for anomaly isolation. On the basis of locating the abnormal block, the contribution plot method is used to locate the abnormal variable to realize anomaly isolation accurately. Suppose that the synthetical dictionary and the analytical coding matrix of the abnormal sample y_K^h are defined as D_K and A_K^h, respectively, the abnormal sample y_K^h can be expressed as follows:

$$y_K^h = D_K A_K^h + f = [D_K, I] \begin{bmatrix} A_K^h \\ f \end{bmatrix} \tag{34}$$

where $I \in R^{s \times s}$ is an identity matrix, and s is the number of variables in the Kth sub-block. The non-zero terms of the vector f represent the position and size of the anomaly source. To more clearly represent the anomaly source, the augmented synthetical dictionary is defined as $\bar{D}_K = [D_K, I]$, so the new analytical coding matrix of the anomaly sample y_K^h under $\bar{D}_K$ is calculated as follows:

$$A_{K,new}^h = \arg\min_{A_K^h} \left\| y_K^h - \bar{D}_K A_K^h \right\|_F^2 \tag{35}$$

Then, the abnormal sample is reformulated as $y_K^h = \bar{D}_K A_{K,new}^h$. In addition, the vector f can be replaced by $SA_{K,new}^h$, where $S = [O, I] \in R^{s \times (d+s)}$, and O is the zero matrix. The variable contribution (VC) of the jth variable in the Kth block is defined by contribution plot method, which can be calculated as follows:

$$VC_K^j = \sum_{h=1}^{H} \left[e_j S A_{K,new}^h \right]^2 \tag{36}$$

And the corresponding variable contribution index (VCI) is expressed as

$$VCI_K^j = \frac{VC_K^j}{\sum_{i=1}^{s} VC_K^i} \tag{37}$$

where $e_j = [\underbrace{0, \cdots, 0}_{j-1}, 1 \underbrace{0, \cdots, 0}_{s-j}]$ is an identity matrix.

3. Experiments

To verify the effectiveness of the proposed method in industrial process monitoring, a numerical simulation experiment is designed and Tennessee Eastman (TE) bench-

mark tests are carried out, and then the proposed method is applied in a real-world aluminum electrolysis industrial process. Besides, the proposed method is compared with several common methods, including robust PCA (rPCA) [35], distributed PCA (DPCA) [36], KSVD [16], DDL [27] and DPL [18]. Meanwhile, Training Time, Testing Time, false alarm rate (FAR), and fault detection rate (FDR) are considered to quantitatively evaluate the performance of different process monitoring methods [24]. To ensure that the comparison is fair, we use the original codes of the comparison methods directly. The number of dictionary atoms is set to 50 for all dictionary learning methods. For the proposed PRDPL method, the optimal parameters of α, β and λ are selected from the candidate set $\left\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\right\}$. With each group of parameters, twenty groups of samples are randomly collected to test. The parameter value corresponding to the highest average FDR is recorded as the parameter setting value. For all comparison methods, we obtain the optimal parameters from the original paper proposing the comparison methods, or adopt the same setting strategy of optimal parameters as our proposed method. Besides, when the comparison method adopts the optimal parameters from the original paper, we also randomly collect 20 groups of samples for the experiment, and take the average value as the experimental result. Meanwhile, the kernel density estimation method is used to obtain the control thresholds of all dictionary learning methods in this paper, where all parameters are set consistently. For rPCA and DPCA, we use the method provided in the original paper to obtain the control threshold. Moreover, the significance level is set to 0.05, which is common to all methods.

3.1. Numerical Simulation Experiment

Firstly, to verify the effectiveness of the proposed method, a linear system for generating high-dimensional data is introduced as follows [27]:

$$X = As + e \tag{38}$$

where $A \in R^{8\times 2}$ represents a random observation matrix, s is a state vector containing two independent variables, and e is a noise vector composed of eight independent Gaussian noises with zero mean and the standard deviation of 0.01. Four different state vectors are designed to simulate different operation units in the process, which is used as prior process knowledge to divide into four sub-blocks. Thus, the state vectors are expressed as follows:

$$\begin{array}{lll} \textbf{\textit{Block}1} & s_1 : U(2,3) & s_2 : N(7,1) \\ \textbf{\textit{Block}2} & s_1 : 2\cos(0.08t)\cdot\sin(0.006t) & s_2 : N(2,0.1) \\ \textbf{\textit{Block}3} & s_1 : 2\cos(0.08t)\cdot\sin(0.006t) & s_2 : U(-1,1) \\ \textbf{\textit{Block}4} & s_1 : U(-1,1) & s_2 : N(2,0.1) \end{array}$$

Then, the aforementioned system is used to generate 32-dimensional process data, and 2000 data are collected as training samples and 300 data are collected as normal testing samples. In addition, a bias fault of 2 is added to the third dimension of the first block, and 500 data are collected as abnormal testing samples. First, we divide the training samples in dimension according to the generation relation, that is, the 32-dimensional data are divided into four sub-blocks of 8-dimensional data. Then, RDPL sub-models are built for each sub-block to learn robust synthetical dictionary and robust analytical dictionary, see robust dictionary pair learning in Algorithm 1. Next, the reconstruction errors are calculated to obtain the control threshold. Finally, we implement anomaly detection and anomaly isolation by GAI defined in Equation (30), BCI defined in Equation (33), and VCI defined in Equation (37). For the diagram of the above method, refer to Figure 2.

The process monitoring results of all methods in the numerical simulation experiment are shown in Figure 5, and the quantitative monitoring results of each method are shown in Table 1. As can be seen from the experimental results, The FDRs of DDL method, DPL method, and DRDPL method all reached 100%. Meanwhile, DRDPL method performs best in process monitoring, and its FAR is as low as 0. The FDRs of T^2 statistics in rPCA

method and DPCA method are up to 100%, but their SPE statistics have a lower FDR in process monitoring. In addition, the training time of rPCA method, DPL method, and DRDPL method is obviously shorter. Although DDL method has better accuracy in process monitoring, it consumes the most computational time.

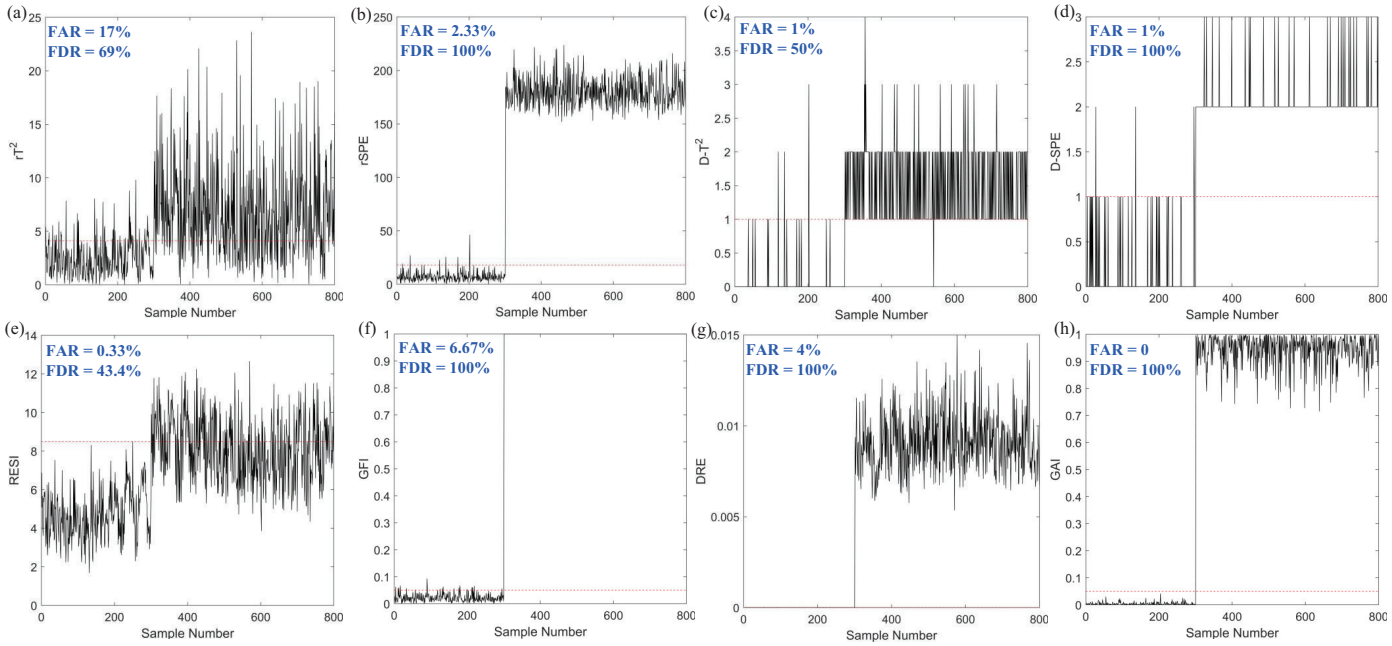

Figure 5. The process monitoring results of numerical simulation experiment. (**a**) rPCA-rT^2; (**b**) rPCA-$rSPE$; (**c**) DPCA-$D - T^2$; (**d**) DPCA-$D - SPE$; (**e**) KSVD-$RESI$; (**f**) DDL-GFI; (**g**) DPL-DRE; (**h**) DRDPL-GAI.

Table 1. The comparison results of numerical simulation experiment.

Method	Training Time (s)	Testing Time (s)	FAR (%)	FDR (%)
rPCA (rT^2)	0.0114	0.0003	17.00	69.00
rPCA ($rSPE$)	0.0112	0.0002	2.33	100.00
DPCA ($D - T^2$)	2.3512	0.0811	1.00	50.00
DPCA ($D - SPE$)	2.7365	0.1406	1.00	100.00
K-SVD ($RESI$)	2.6444	0.0039	0.33	43.40
DDL (GFI)	10.5511	0.0036	6.67	100.00
DPL (DRE)	0.0316	0.4884	4.00	100.00
DRDPL (*GAI*)	**0.2702**	**0.0033**	**0.00**	**100.00**

The anomaly isolation results of numerical simulation experiment are shown in Figure 6. Figure 6a shows the result of locating the abnormal block by BCI, which demonstrates that the anomaly occurred in the first block. On the basis of locating abnormal blocks, Figure 6b shows that the abnormal variable is further located by VCI. The location result demonstrates that the anomaly is most likely to occur in the third variable. Therefore, the results of anomaly location ate consistent with the anomaly we set.

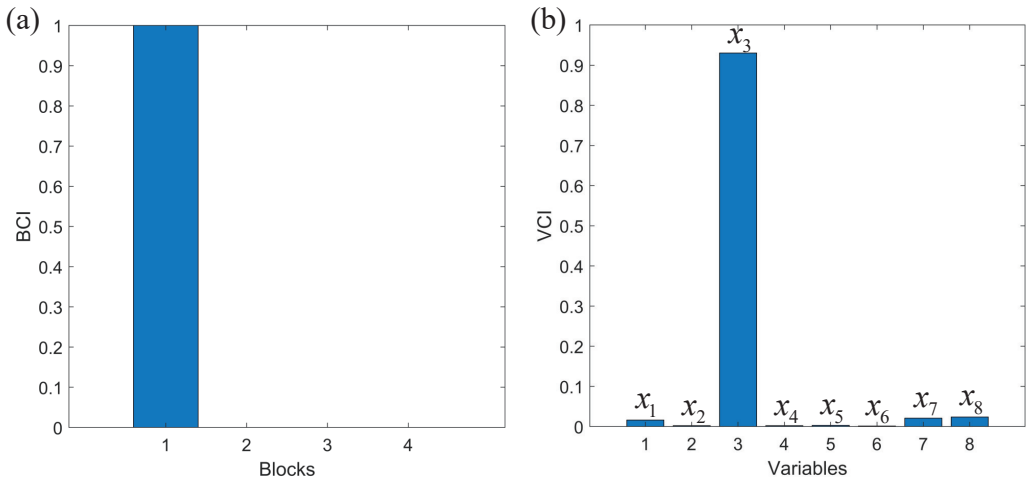

Figure 6. The anomaly isolation results in numerical simulation experiment. (**a**) BCI; (**b**) VCI.

3.2. TE Benchmark Test

Tennessee Eastman (TE) benchmark tests are often used to validate process monitoring methods. The structure diagram of TE process is shown in Figure 7. There are mainly 5 operation units, including 12 process control variables and 41 process measurement variables [37]. Notably, 22 process measurement variables and 9 process control variables are selected as 31 process variables in TE benchmark tests, see Table A1 in Appendix A. In addition, according to the technological process of TE process, 31 variables can be divided into four blocks, which are shown in Table 2 [25].

Table 2. The division of TE process variables.

Block	Variables	Principle of Division
1	$v_1, v_2, v_3, v_5, v_6, v_{23}, v_{24}, v_{25}$	Input
2	$v_7, v_8, v_9, v_{21}, v_{30}$	Reactor
3	$v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{20}, v_{22}, v_{27}, v_{28}, v_{31}$	Separator, Compressor and Condenser
4	$v_4, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{26}, v_{29}$	Stripper

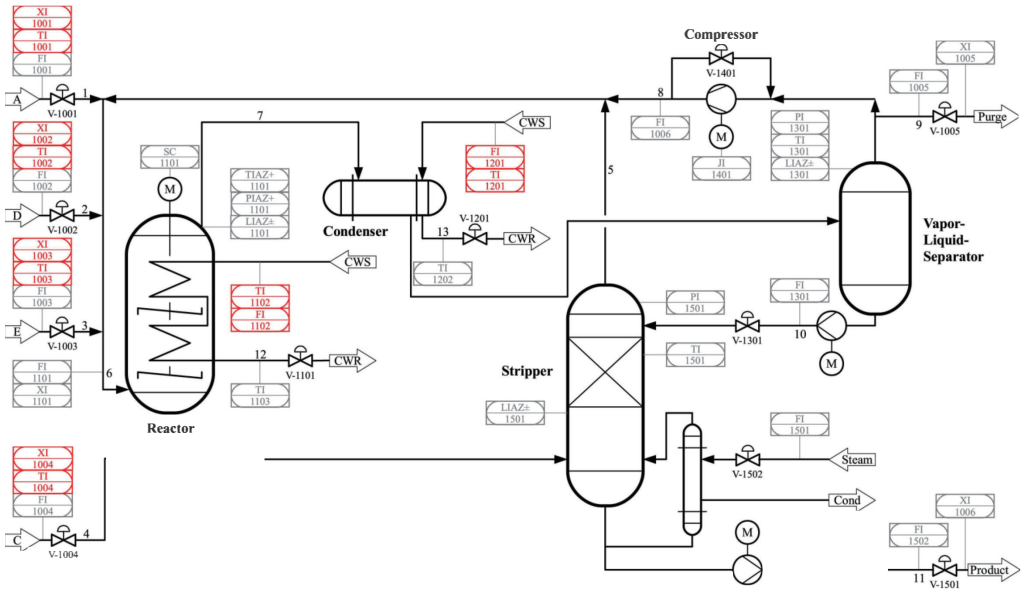

Figure 7. Structural schematic diagram of TE process.

TE process contains 28 disturbances, see Table A2 in Appendix A. 3000 data are collected as training samples and 300 data are collected as normal testing samples. Moreover,

500 data are collected under each disturbance as abnormal testing samples. Similarly, for the schematic diagram of TE process monitoring based on DRDPL, refer to Figure 2. Table 3 shows that the FDRs of all methods in TE processes with various disturbances. The results of process monitoring show that the proposed method has high FDR for most disturbances.

Table 3. The FDR of the proposed method and comparative methods for 28 kinds of disturbances in TE process.

Disturbance Number	rPCA (rT^2)	rPCA ($rSPE$)	DPCA ($D-T^2$)	DPCA ($D-SPE$)	K-SVD ($RESI$)	DDL (GFI)	DPL (DRE)	DRDPL (GAI)
1	0.9900	0.9900	1.0000	1.0000	0.9760	0.9880	0.9460	**1.0000**
2	0.9640	0.9660	1.0000	1.0000	0.9300	0.9560	0.9040	**1.0000**
3	0.9900	0.9880	1.0000	1.0000	0.9740	0.9840	0.9460	**1.0000**
4	0.9980	0.9980	1.0000	0.4260	0.9980	0.9980	0.0700	**1.0000**
5	0.9460	0.9820	1.0000	1.0000	0.9100	0.9660	0.8920	**1.0000**
6	0.9980	0.9980	0.9980	0.9980	0.9980	0.9980	0.9920	**0.9980**
7	0.9980	0.9980	1.0000	1.0000	0.9980	0.9980	0.9980	**1.0000**
8	0.8020	0.8320	1.0000	1.0000	0.7360	0.7960	0.7180	**1.0000**
9	0.9980	0.9980	1.0000	0.4000	0.9980	0.9980	0.1000	**1.0000**
10	0.7720	0.9260	0.9760	0.9660	0.5600	0.8700	0.0320	**0.8000**
11	0.9500	0.9420	1.0000	0.9760	0.9160	0.9180	0.3420	**0.9880**
12	0.5620	0.4420	0.9300	0.8420	0.2900	0.4340	0.0540	**0.1820**
13	0.9500	0.9460	1.0000	1.0000	0.9280	0.9360	0.9120	**1.0000**
14	0.9633	0.8000	1.0000	1.0000	0.9540	0.9680	0.0260	**0.9940**
15	0.9860	0.9880	1.0000	1.0000	0.9620	0.9600	0.0620	**0.9880**
16	0.9500	0.9460	1.0000	1.0000	0.9260	0.9340	0.9120	**1.0000**
17	0.9600	0.9500	1.0000	1.0000	0.9320	0.9320	0.4180	**0.9320**
18	0.7060	0.7720	0.9560	0.9860	0.5740	0.7520	0.2680	**0.8120**
19	0.9780	0.9540	1.0000	1.0000	0.8760	0.9620	0.0460	**1.0000**
20	0.8340	0.7940	1.0000	1.0000	0.7680	0.7880	0.7360	**0.9020**
21	0.8320	0.7920	1.0000	1.0000	0.7720	0.7860	0.7360	**0.9300**
22	0.9600	0.9360	1.0000	1.0000	0.8420	0.9460	0.0480	**0.9540**
23	0.7200	0.7840	0.9760	0.9960	0.5900	0.8600	0.2900	**0.7360**
24	0.8360	0.9160	1.0000	1.0000	0.7780	0.9300	0.4260	**0.9220**
25	0.7760	0.9360	0.9880	0.9800	0.5800	0.9180	0.2120	**0.7040**
26	0.9560	0.9600	1.0000	0.9960	0.9260	0.9600	0.2920	**0.9500**
27	0.9320	0.9580	0.9820	0.6300	0.8240	0.8900	0.0660	**0.8280**
28	0.8440	0.9100	0.9900	0.8120	0.6180	0.8260	0.0280	**0.8220**

To further compare the process monitoring performance of the proposed method with other methods, the results of TE processes under disturbance IDV (6) and IDV (14) are shown in Figures 8 and 9, respectively, and the quantitative monitoring results are presented in Tables 4 and 5, respectively. The experimental results show that the proposed DRDPL method performs better in process monitoring. That is, rPCA method, DPCA method, K-SVD method, DDL method, and the proposed DRDPL method have high FDRs, while the FAR of our DRDPL is the lowest. Meanwhile, the training time of DPCA method and DDL method are significantly more than that of rPCA method, K-SVD method, DPL method, and the proposed DRDPL method. Besides, DPL method has the excellent performance of process monitoring in TE process with IDV (6), but its FDR in TE process with IDV (14) is only 2.60%. Thus, the stability in process monitoring of DPL needs to be improved.

Table 4. The comparison results of TE process with IDV (6).

Method	Training Time (s)	Testing Time (s)	FAR (%)	FDR (%)
rPCA (rT^2)	2.5695	0.0008	16.00	99.80
rPCA ($rSPE$)	1.0328	0.0007	7.33	99.80
DPCA ($D-T^2$)	14.9032	0.7141	22.67	99.80
DPCA ($D-SPE$)	14.8002	0.7019	17.00	99.80
K-SVD ($RESI$)	4.7873	0.0028	2.33	99.80
DDL (GFI)	13.6185	0.0001	6.67	99.80
DPL (DRE)	0.0580	0.1415	4.00	99.20
DRDPL (GAI)	**2.9189**	**0.0045**	**0.00**	**99.80**

Table 5. The comparison results of TE process with IDV (14).

Method	Training Time (s)	Testing Time (s)	FAR (%)	FDR (%)
rPCA (rT^2)	1.2971	0.0002	19.00	99.00
rPCA ($rSPE$)	0.0094	0.0001	4.33	99.40
DPCA ($D-T^2$)	14.3749	0.6949	22.67	100.00
DPCA ($D-SPE$)	14.2242	0.6752	17.00	100.00
K-SVD ($RESI$)	4.8615	0.0029	0.67	95.40
DDL (GFI)	13.8954	0.0001	6.00	96.80
DPL (DRE)	0.0518	0.1417	4.00	2.60
DRDPL (GAI)	**1.3678**	**0.0011**	**0.00**	**99.40**

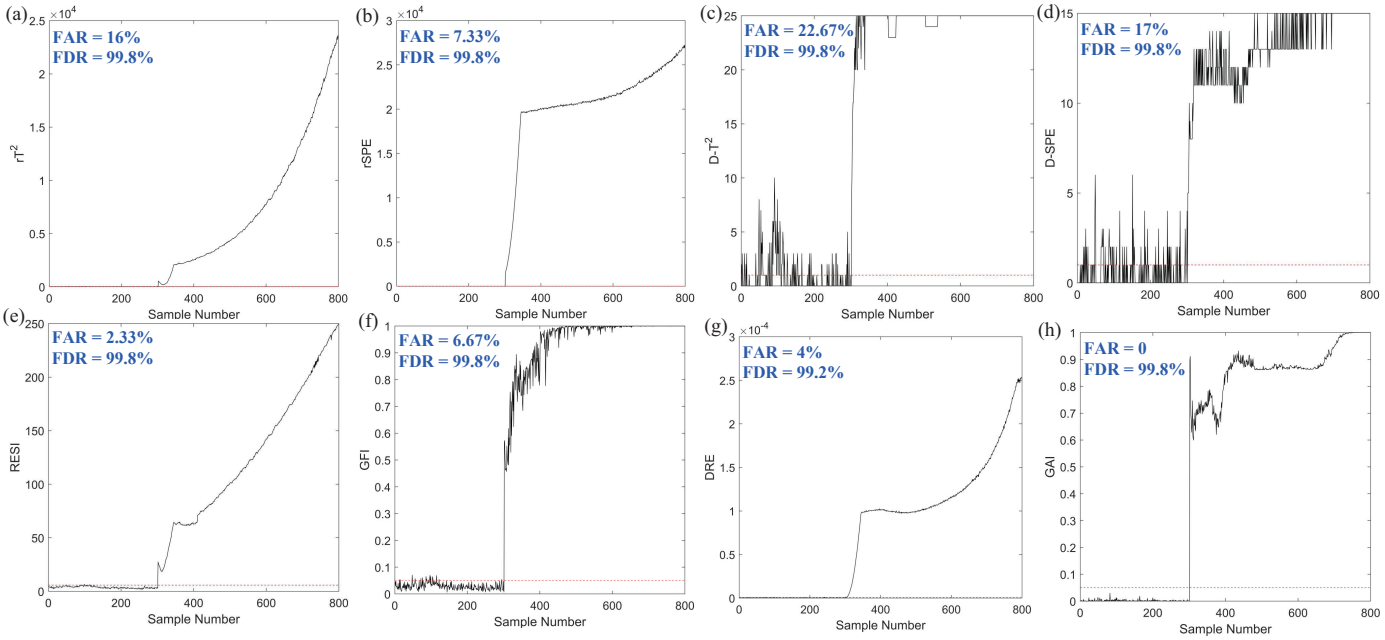

Figure 8. The process monitoring results of TE process with IDV (6). (**a**) rPCA-rT^2; (**b**) rPCA-$rSPE$; (**c**) DPCA-$D-T^2$; (**d**) DPCA-$D-SPE$; (**e**) KSVD-$RESI$; (**f**) DDL-GFI; (**g**) DPL-DRE; (**h**) DRDPL-GAI.

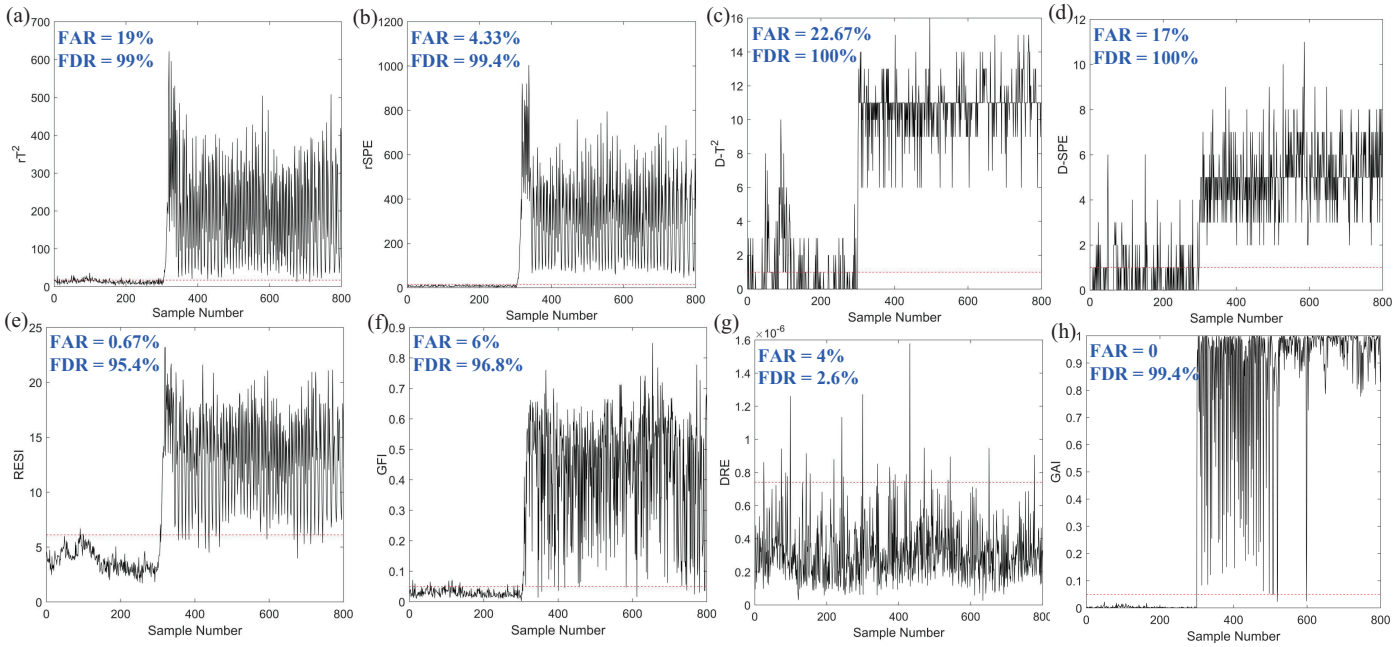

Figure 9. The process monitoring results of TE process with IDV (14). (**a**) rPCA-rT^2; (**b**) rPCA-$rSPE$; (**c**) DPCA-$D-T^2$; (**d**) DPCA-$D-SPE$; (**e**) KSVD-$RESI$; (**f**) DDL-GFI; (**g**) DPL-DRE; (**h**) DRDPL-GAI.

Figures 10 and 11 show the results of anomaly isolation under the two disturbances, respectively. Figures 10a and 11a show the results of locating abnormal blocks by BCI, and Figures 10b and 11b show the results of locating abnormal variables by VCI in sub-blocks with the highest probability of anomalies. The location results show that the first dimension (v_1) of the first block is the most likely to have an anomaly in TE process monitoring with the disturbance of IDV (6), and the fifth dimension (v_{30}) of the second block is most likely to have an anomaly in TE process monitoring with the disturbance of IDV (14). In addition, for the results of variable isolation in other blocks of TE process, see Figure A1 in Appendix B. The results of anomaly location are basically consistent with the possible caused results of disturbances.

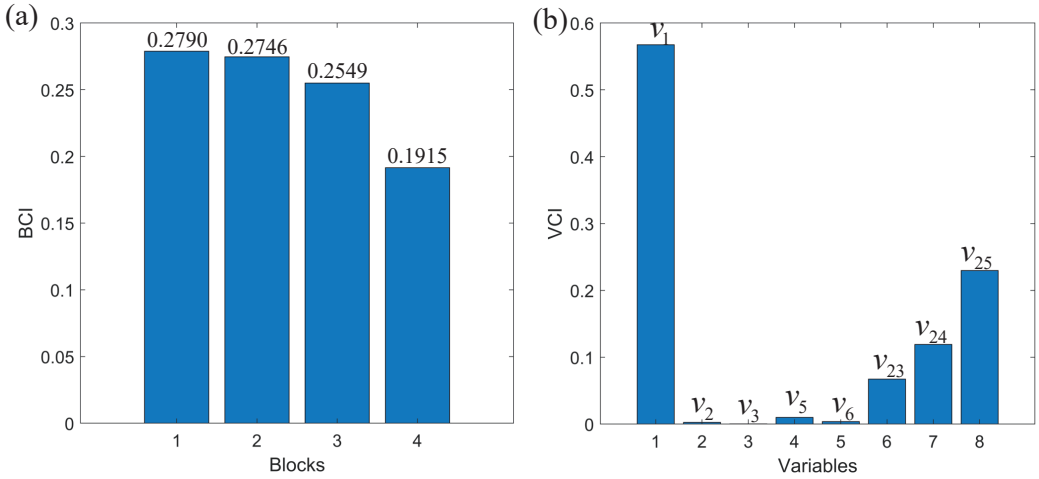

Figure 10. The anomaly isolation results in TE process with IDV (6). (**a**) BCI; (**b**) VCI-Block 1.

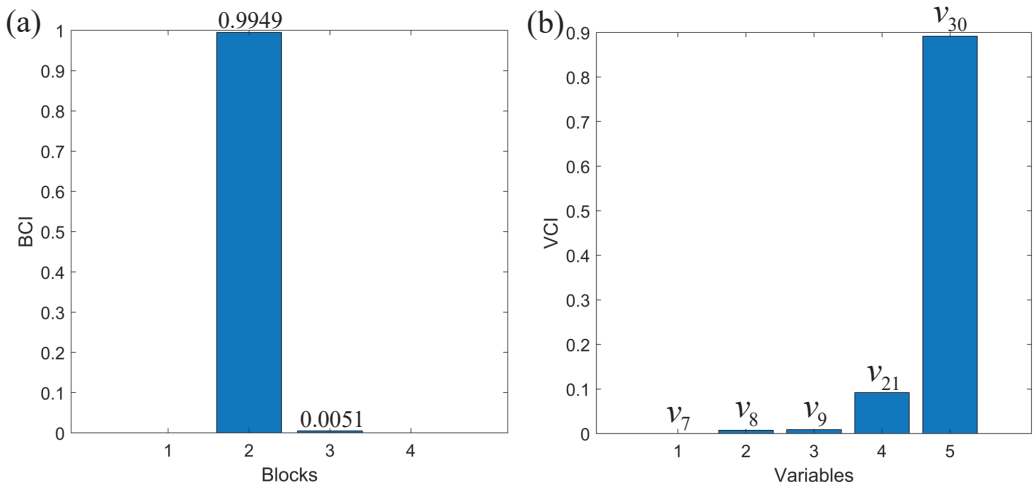

Figure 11. The anomaly isolation results in TE process with IDV (14). (**a**) BCI; (**b**) VCI-Block 2.

3.3. Aluminum Electrolysis Industrial Process Application

With the development of large-scale aluminum electrolytic cell, the complexity and uncertainty of aluminum electrolytic system become higher. Effective cell condition monitoring technology plays an important role in detecting and predicting abnormal cell condition in real time, adjusting control strategy in time, and ensuring efficient and high-quality operation of electrolytic cell [38]. In aluminum electrolysis industrial process, the dynamic behaviors of local anodes form the distributed cell state, and anode current can reflect localized cell states.

In the experiments, anode current data are obtained by the production data report of 400 kA series electrolytic cell from an aluminum electrolysis factory in Shandong Province. The structure diagram of an aluminum electrolysis cell is shown in Figure 12. The series of electrolytic cells have 12 anodes on each side, that is, aluminum electrolysis cells can produce 24-dimensional anode current data. Importantly, six adjacent anodes are divided into one node based on prior process knowledge of alumina concentration in aluminum electrolysis process. Therefore, the 24-dimensional anode current data are divided into four sub-blocks. Meanwhile, 3000 data are collected as training samples and 300 data are collected as normal testing samples. In addition, 500 data are collected under the anodic effect and anodic slip conditions as abnormal test data, respectively. Similarly, for the schematic diagram of aluminum electrolysis industrial process monitoring based on DRDPL, refer to Figure 2.

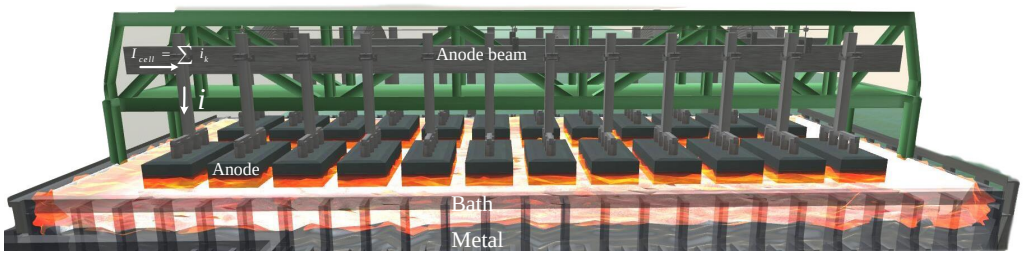

Figure 12. Structural schematic diagram of aluminum electrolysis cell.

Figures 13 and 14 show the monitoring results of aluminum electrolysis processes with the abnormal cell conditions of anode effect and anode slippage. Tables 6 and 7 give corresponding quantitative monitoring results, respectively. The experimental results show that rPCA method, DPL method, and the proposed DRDPL method have outstanding advantages in computional time. In terms of process monitoring performance, the FDRs of

all methods can reach 100%. And the FARs of KSVD method and our DRDPL method are lower than that of rPCA method, DPCA method, DDL method, and DPL method, which are only 0. Therefore, the performance of the proposed method is better than other methods in cell condition monitoring.

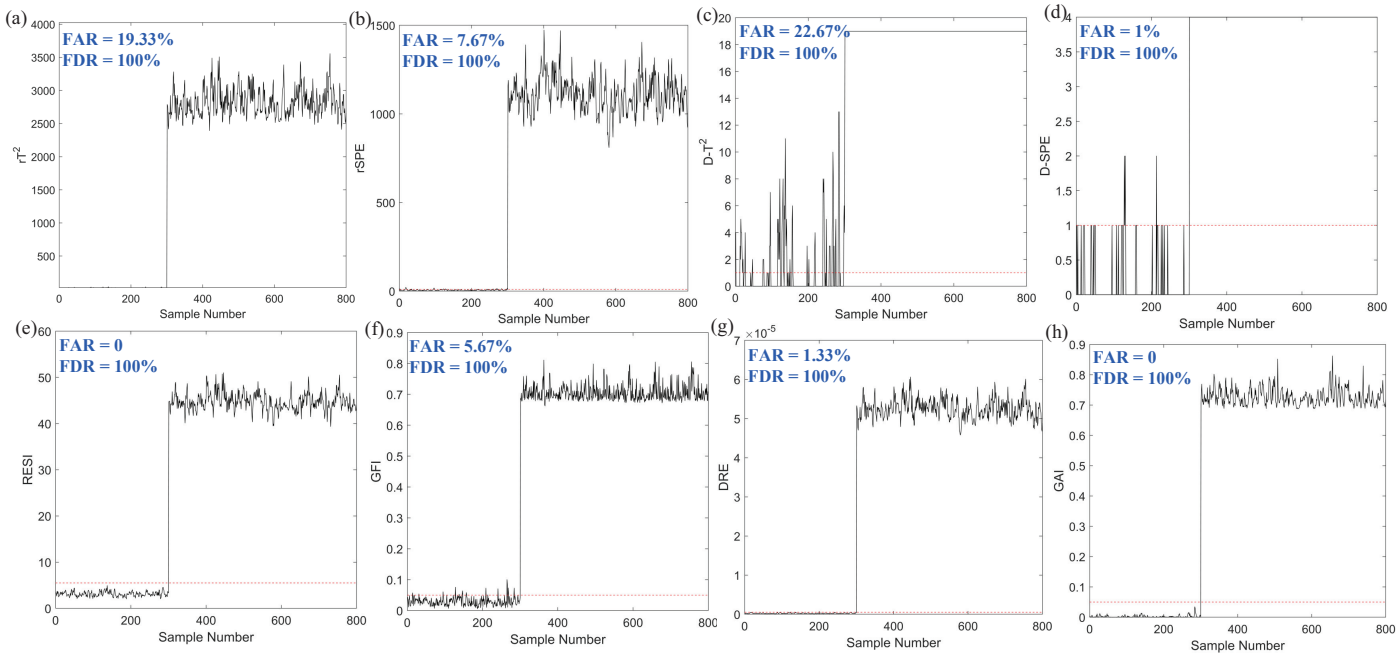

Figure 13. The process monitoring results of anode effect in aluminum electrolysis process. (**a**) rPCA-rT^2; (**b**) rPCA-$rSPE$; (**c**) DPCA-$D-T^2$; (**d**) DPCA-$D-SPE$; (**e**) KSVD-$RESI$; (**f**) DDL-GFI; (**g**) DPL-DRE; (**h**) DRDPL-GAI.

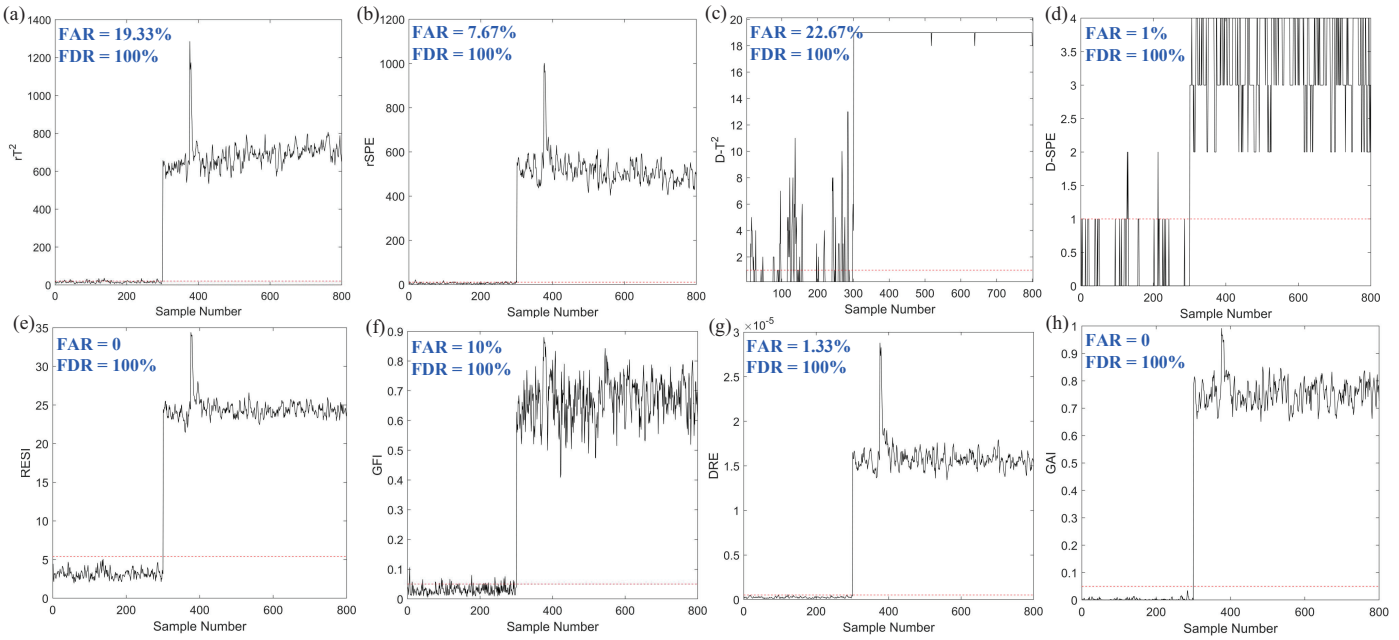

Figure 14. The process monitoring results of anode slippage in aluminum electrolysis process. (**a**) rPCA-rT^2; (**b**) rPCA-$rSPE$; (**c**) DPCA-$D-T^2$; (**d**) DPCA-$D-SPE$; (**e**) KSVD-$RESI$; (**f**) DDL-GFI; (**g**) DPL-DRE; (**h**) DRDPL-GAI.

Table 6. The comparison results of anode effect in aluminum electrolysis process.

Method	Training Time (s)	Testing Time (s)	FAR (%)	FDR (%)
rPCA (rT^2)	0.0152	0.0003	19.33	100.00
rPCA ($rSPE$)	0.0145	0.0001	7.67	100.00
DPCA ($D-T^2$)	13.3106	0.5665	22.67	100.00
DPCA ($D-SPE$)	13.2324	0.5647	1.00	100.00
K-SVD ($RESI$)	3.6734	0.0049	0.00	100.00
DDL (GFI)	12.8322	0.0004	5.67	100.00
DPL (DRE)	0.0662	0.2882	1.33	100.00
DRDPL (GAI)	**0.6024**	**0.0012**	**0.00**	**100.00**

Table 7. The comparison results of anode slippage in aluminum electrolysis process.

Method	Training Time (s)	Testing Time (s)	FAR (%)	FDR (%)
rPCA (rT^2)	0.0200	0.0008	19.33	100.00
rPCA ($rSPE$)	0.0192	0.0007	7.67	100.00
DPCA ($D-T^2$)	13.2074	0.5386	22.67	100.00
DPCA ($D-SPE$)	13.8177	0.5718	1.00	100.00
K-SVD ($RESI$)	4.0545	0.0053	0.00	100.00
DDL (GFI)	13.0848	0.0024	10.00	100.00
DPL (DRE)	0.0649	0.2841	1.33	100.00
DRDPL (GAI)	**0.6090**	**0.0014**	**0.00**	**100.00**

Figures 15 and 16 show the anomaly isolation results of occurring anode effect and anode slippage, respectively. Figures 15a and 16a show the results of locating abnormal blocks by BCI, and Figures 15b and 16b show the results of locating abnormal variables by VCI in sub-blocks with the highest probability of anomalies. The location results show that the fourth dimension (i_{13}) of the first block is the most likely to have an anomaly in aluminum electrolysis processes with anode effect and anode slippage, which is consistent with the anomaly source location of the actual cell condition. In addition, for the results of variable isolation in other blocks of aluminum electrolysis industrial process, see Figures A2 and A3 in Appendix B.

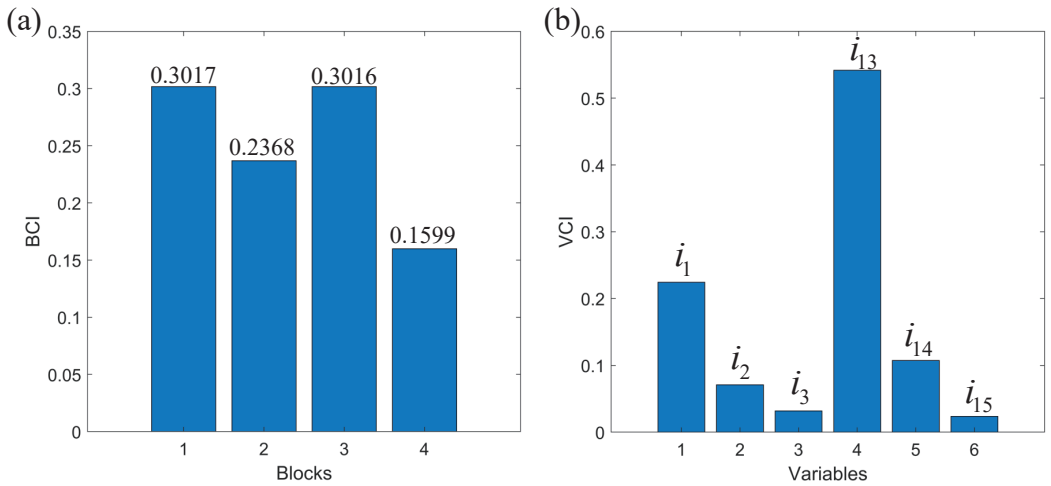

Figure 15. The anomaly isolation results of anode effect in aluminum electrolysis process. (**a**) BCI; (**b**) VCI-Block 1.

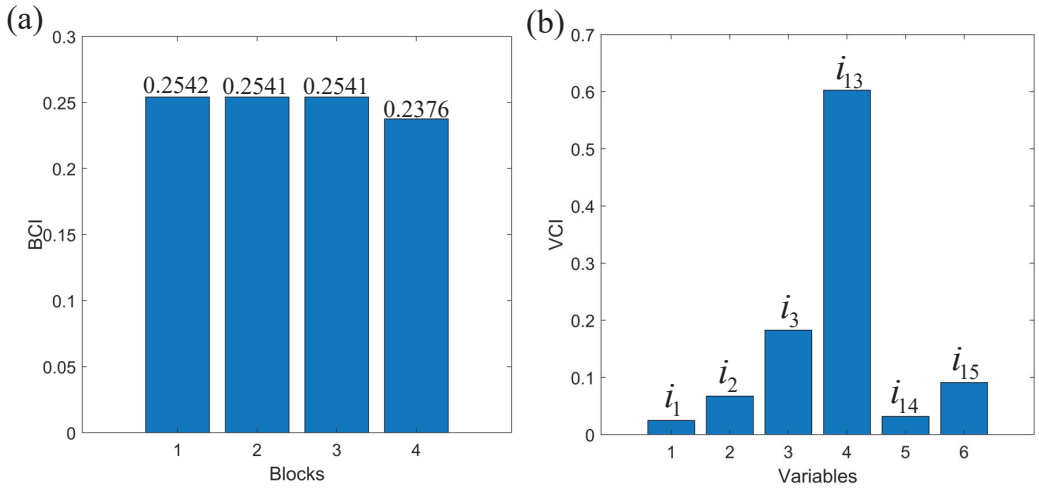

Figure 16. The anomaly isolation results of anode slippage in aluminum electrolysis process. (**a**) BCI; (**b**) VCI-Block 1.

3.4. Fault Detection against Noisy Datasets

To test the robustness of the proposed DRDPL method, we add random Gaussian noise to datasets by $Data = Data + \sqrt{Variance} \times randn(size(Data))$. For the numerical simulation experiment, TE process experiment with IDV (6), and aluminum electrolysis process experiment with anode effect, 20 groups of samples are randomly selected for testing, and average values are taken as the results of fault detection. The variance values are set in ranges of $\{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$, $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, and $\{1.1 \times 10^5, 1.2 \times 10^5, 1.3 \times 10^5, 1.4 \times 10^5, 1.5 \times 10^5, 1.6 \times 10^5, 1.7 \times 10^5, 1.8 \times 10^5, 1.9 \times 10^5, 2.0 \times 10^5\}$, respectively. Figure 17 shows the experimental results in the case of noise. We can find that the overall trend of FDR for each method decreases with the increase of variance. It is worth noting that our proposed DRDPL method provides higher FDR than other methods in most cases. That is, the DRDPL method is more robust to the interference of noise due to the adoption of a more reasonable mechanism.

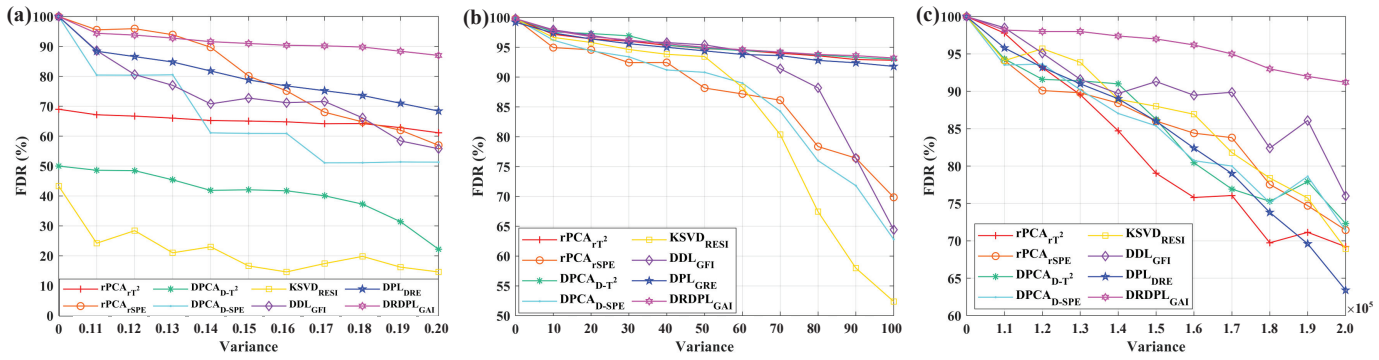

Figure 17. FDR of each method with varying variance. (**a**) Numerical simulation; (**b**) TE process with IDV (6); (**c**) Aluminum electrolysis process with anode effect.

4. Conclusions

In this article, a process monitoring method based on distributed robust dictionary pair learning (PRDPL) is proposed for anomaly detection and anomaly isolation. Firstly, the reliable prior knowledge of industrial processes is integrated into the data-driven model with block division, which is conducive to exposing small anomalies in high-dimensional data. Then, a robust dictionary pair learning (RDPL) method is proposed to build a local monitoring model for each sub-block to obtain robust dictionary pairs. Finally, Bayesian inference method is introduced to realize anomaly detection. To further find the anomaly sources, the block contribution index and variable contribution index are defined to locate

abnormal blocks and abnormal variables, respectively. Thus, the applicability and reliability of the proposed method have been demonstrated in a numerical simulation, Tennessee Eastman processes and aluminum electrolysis processes. Particularly, our method performs well in anomaly detection, computation time, and robustness together. It is worth noted that the prior process knowledge used for division has not been systematized. Therefore, a more accurate block division method based on the fusion of knowledge and data is worth further study. In addition, how to obtain adaptive dictionary learning models by considering abnormal samples is a meaningful research direction.

Author Contributions: Methodology, software, writing—original draft, J.W., X.C. and Z.D.; writing—review and editing, Z.D.; formal analysis, supervision, funding acquisition, conceptualization, X.C. and H.Z.; data curation, validation, visualization, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the State Key Program of the National Natural Science of China under Grant 62133016, in part by the National Natural Science Foundation of China under Grants 51974373 and 51874365, and in part by the Yunnan Province Science and Technology Planning Project under Grants 202202AB080017 and 202102AB080062.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The introduction of each variable in TE benchmark test is shown in Table A1. The descriptions of 28 kinds of disturbances in TE process are shown in Table A2.

Table A1. The process variables of TE benchmark test.

No.	Variables	No.	Variables
1	A feed (stream 1)	17	Stripper underflow (stream 11)
2	D feed (stream 2)	18	Stripper temperature
3	E feed (stream 3)	19	Stripper steam flow
4	A and C feed (stream 4)	20	Compressor work
5	Recycle flow (stream 8)	21	Reactor cooling water outlet temperature
6	Reactor feed rate (stream 6)	22	Separator cooling water outlet temperature
7	Reactor pressure	23	D feed flow (stream 2)
8	Reactor level	24	E feed flow (stream 3)
9	Reactor temperature	25	A feed flow (stream 1)
10	Purge rate (stream 9)	26	A and C feed flow (stream 4)
11	Product separator temperature	27	Purge valve (stream 9)
12	Product separator level	28	Separator pot liquid flow (stream 10)
13	Product separator pressure	29	Stripper liquid product flow (stream 11)
14	Product separator underflow (stream 10)	30	Reactor cooling water flow
15	Stripper level	31	Stripper liquid product flow (stream 11)
16	Stripper pressure		

Table A2. The disturbances of TE process.

Number	Disturbance Description	Type
IDV (1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV (2)	B composition, A/C ratio constant (stream 4)	Step
IDV (3)	D feed temperature (stream 2)	Step
IDV (4)	Reactor cooling water inlet temperature	Step
IDV (5)	Condenser cooling water inlet temperature	Step
IDV (6)	A feed loss (stream 1)	Step
IDV (7)	C header pressure loss-reduced availability (stream 4)	Step
IDV (8)	A, B, C feed composition (stream 4)	Random variation
IDV (9)	D feed temperature (stream 2)	D feed temperature
IDV (10)	C feed temperature (stream 4)	Random variation
IDV (11)	Reactor cooling water inlet temperature	Random variation
IDV (12)	Condenser cooling water inlet temperature	Random variation
IDV (13)	Reaction kinetics	Slow drift
IDV (14)	Reactor cooling water valve	Sticking
IDV (15)	Condenser cooling water valve	Sticking
IDV (16)	Unknown	Unknown
IDV (17)	Unknown	Unknown
IDV (18)	Unknown	Unknown
IDV (19)	Unknown	Unknown
IDV (20)	Unknown	Unknown
IDV (21)	A feed temperature (stream 1)	-
IDV (22)	E feed temperature (stream 3)	-
IDV (23)	A feed pressure (stream 1)	-
IDV (24)	D feed pressure (stream 2)	-
IDV (25)	E feed pressure (stream 3)	-
IDV (26)	A and C feed pressure (stream 4)	-
IDV (27)	Pressure fluctuation in the cooling water re-circulating unit of the reactor	-
IDV (28)	Pressure fluctuation in the cooling water re-circulating unit of the condenser	-

Appendix B

In the actual industrial production process, the purpose of anomaly isolation is to find the block with the highest probability of anomalies, so as to provide the basis for process control and decision. Therefore, the variable isolation results of TE process and aluminum electrolysis industrial process except for the most likely abnormal blocks are shown in Figures A1, A2 and A3, respectively.

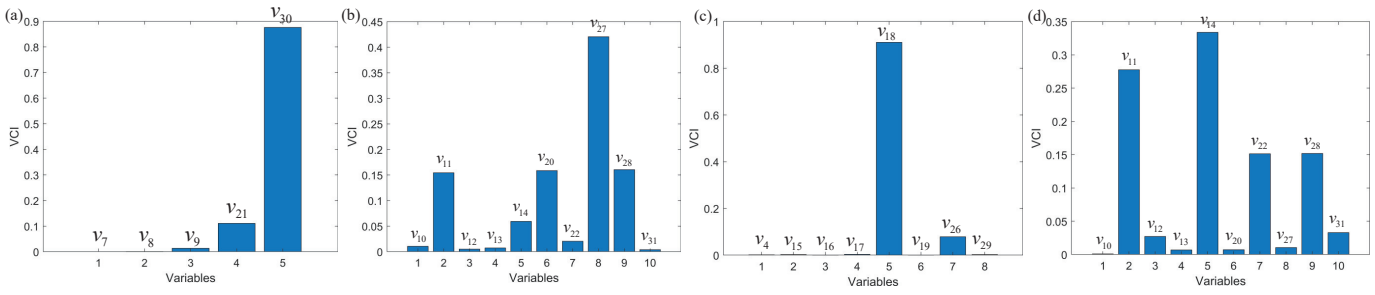

Figure A1. The variable isolation results in TE process. (**a**) Block 2 with IDV (6); (**b**) Block 3 with IDV (6); (**c**) Block 4 with IDV (6); (**d**) Block 3 with IDV (14).

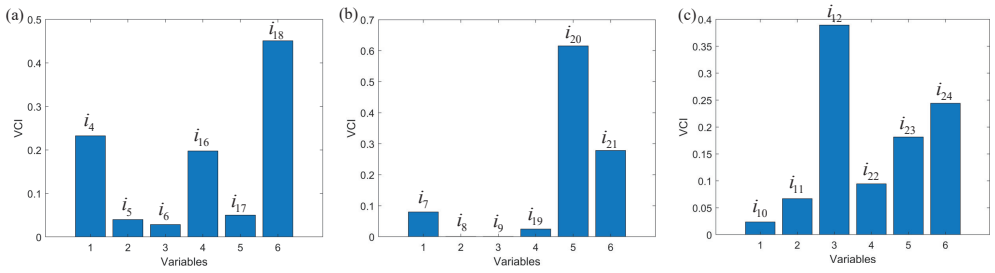

Figure A2. The variable isolation results in anode effect of aluminum electrolysis industrial process. (**a**) Block 2; (**b**) Block 3; (**c**) Block 4.

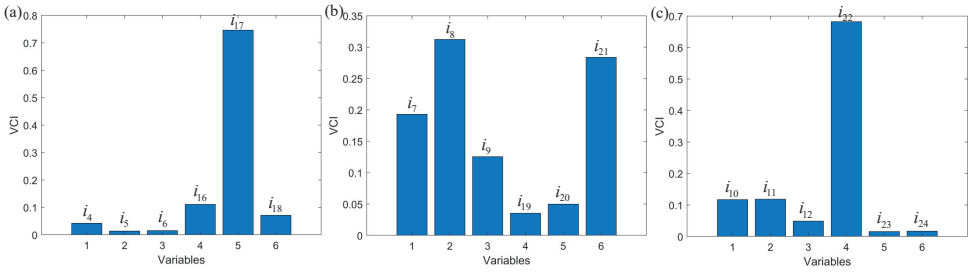

Figure A3. The variable isolation results in anode slippage of aluminum electrolysis industrial process. (**a**) Block 2; (**b**) Block 3; (**c**) Block 4.

References

1. Yang, F.; Cui, Y.; Wu, F.; Zhang, R. Fault Monitoring of Chemical Process Based on Sliding Window Wavelet DenoisingGLPP. *Processes* **2021**, *9*, 86. [CrossRef]
2. Wang, J.; Zhou, Z.; Li, Z.; Du, S. A Novel Fault Detection Scheme Based on Mutual k-Nearest Neighbor Method: Application on the Industrial Processes with Outliers. *Processes* **2022**, *10*, 497. [CrossRef]
3. Zhang, H.; Li, T.; Li, J.; Yang, S.; Zou, Z. Progress in aluminum electrolysis control and future direction for smart aluminum electrolysis plant. *JOM* **2017**, *69*, 292–300. [CrossRef]
4. Cheung, C.Y.; Menictas, C.; Bao, J.; Skyllas-Kazacos, M.; Welch, B.J. Characterization of Individual Anode Current Signals in Aluminum Reduction Cells. *Ind. Eng. Chem. Res.* **2013**, *52*, 9632–9644. [CrossRef]
5. Zhai, S.; Wang, W.; Ye, H. Fault diagnosis based on parameter estimation in closed-loop systems. *IET Control Theory Appl.* **2015**, *9*, 1146–1153. [CrossRef]
6. Chi, G.; Wang, D. Sensor Placement for Fault Isolability Based on Bond Graphs. *IEEE Trans. Autom. Control* **2015**, *60*, 3041–3046. [CrossRef]
7. Yue, W.; Gui, W.; Xie, Y. Experiential knowledge representation and reasoning based on linguistic Petri nets with application to aluminum electrolysis cell condition identification. *Inf. Sci.* **2020**, *529*, 141–165. [CrossRef]
8. Tessier, J.; Duchesne, C.; Tarcy, G.P.; Gauthier, C.; Dufour, G. Multivariate Analysis and Monitoring of the Performance of Aluminum Reduction Cells. *Ind. Eng. Chem. Res.* **2012**, *51*, 1311–1323. [CrossRef]
9. Huang, Z.; Yang, C.; Chen, X.; Zhou, X.; Chen, G.; Huang, T.; Gui, W. Functional deep echo state network improved by a bi-level optimization approach for multivariate time series classification. *Appl. Soft Comput.* **2021**, *106*, 107314. [CrossRef]
10. Huang, K.; Wu, Y.; Wang, C.; Xie, Y.; Yang, C.; Gui, W. A Projective and Discriminative Dictionary Learning for High-Dimensional Process Monitoring With Industrial Applications. *IEEE Trans. Ind. Inform.* **2021**, *17*, 558–568. [CrossRef]
11. Ji, C.; Sun, W. A Review on Data-Driven Process Monitoring Methods: Characterization and Mining of Industrial Data. *Processes* **2022**, *10*, 335. [CrossRef]
12. Huang, K.; Wei, K.; Zhou, L.; Li, Y.; Yang, C. Multimode Process Monitoring and Mode Identification Based on Multiple Dictionary Learning. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
13. Deng, Z.; Chen, X.; Xie, S.; Xie, Y.; Zhang, H. Semi-Supervised Discriminative Projective Dictionary Pair Learning and Its Application for Industrial Process Monitoring. *IEEE Trans. Ind. Inform.* **2022**. [CrossRef]
14. Zhang, H.; Chen, X.; Du, Z.; Yan, R. Kurtosis based weighted sparse model with convex optimization technique for bearing fault diagnosis. *Mech. Syst. Signal Process.* **2016**, *80*, 349–376. [CrossRef]
15. Du, Z.; Chen, X.; Zhang, H.; Yan, R. Sparse Feature Identification Based on Union of Redundant Dictionary for Wind Turbine Gearbox Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6594–6605. [CrossRef]
16. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [CrossRef]
17. Jiang, Z.; Lin, Z.; Davis, L.S. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2651–2664. [CrossRef]

18. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Projective dictionary pair learning for pattern classification. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 793–801.
19. Zhang, Z.; Sun, Y.; Wang, Y.; Zhang, Z.; Zhang, H.; Liu, G.; Wang, M. Twin-Incoherent Self-Expressive Locality-Adaptive Latent Dictionary Pair Learning for Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 947–961. [CrossRef]
20. Sun, Y.; Zhang, Z.; Jiang, W.; Zhang, Z.; Zhang, L.; Yan, S.; Wang, M. Discriminative Local Sparse Representation by Robust Adaptive Dictionary Pair Learning. *IEEE Trans. Neural Networks Learn. Syst.* **2020**, *31*, 4303–4317. [CrossRef]
21. Li, N.; Guo, S.; Wang, Y. Weighted preliminary-summation-based principal component analysis for non-Gaussian processes. *Control Eng. Pract.* **2019**, *87*, 122–132. [CrossRef]
22. Liu, Q.; Qin, S.J.; Chai, T. Multiblock Concurrent PLS for Decentralized Monitoring of Continuous Annealing Processes. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6429–6437. [CrossRef]
23. He, F.; Xu, J. A novel process monitoring and fault detection approach based on statistics locality preserving projections. *J. Process Control* **2016**, *37*, 46–57. [CrossRef]
24. Deng, Z.; Chen, X.; Xie, S.; Xie, Y.; Sun, Y. Distributed process monitoring based on joint mutual information and projective dictionary pair learning. *J. Process Control* **2021**, *106*, 130–141. [CrossRef]
25. Zhu, J.; Ge, Z.; Song, Z. Distributed Parallel PCA for Modeling and Monitoring of Large-Scale Plant-Wide Processes With Big Data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1877–1885. [CrossRef]
26. Xu, C.; Liu, F. Process monitoring based on distributed principal component analysis with angle-relevant variable selection. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1–13. [CrossRef]
27. Huang, K.; Wu, Y.; Wen, H.; Liu, Y.; Yang, C.; Gui, W. Distributed dictionary learning for high-dimensional process monitoring. *Control Eng. Pract.* **2020**, *98*, 104386. [CrossRef]
28. Zhang, Z.; Jia, L.; Zhao, M.; Liu, G.; Wang, M.; Yan, S. Kernel-Induced Label Propagation by Mapping for Semi-Supervised Classification. *IEEE Trans. Big Data* **2019**, *5*, 148–165. [CrossRef]
29. Cai, D.; He, X.; Han, J.; Huang, T.S. Graph Regularized Nonnegative Matrix Factorization for Data Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1548–1560.
30. Wang, Y.; Du, H.; Zhang, Y.; Zhang, Y. Efficient and robust discriminant dictionary pair learning for pattern classification. *Digit. Signal Process.* **2021**, *118*, 103227. [CrossRef]
31. Mazumder, R.; Hastie, T.; Tibshirani, R. Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **2010**, *11*, 2287–2322. [PubMed]
32. Terrell, G.R.; Scott, D.W. Variable kernel density estimation. *Ann. Stat.* **1992**, *20*, 1236–1265. [CrossRef]
33. Xu, C.; Zhao, S.; Liu, F. Distributed plant-wide process monitoring based on PCA with minimal redundancy maximal relevance. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 53–63. [CrossRef]
34. Jiang, Q.; Yan, X. Monitoring multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability, and Bayesian inference. *Chemom. Intell. Lab. Syst.* **2014**, *136*, 121–137. [CrossRef]
35. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM* **2011**, *58*, 1–37. [CrossRef]
36. Ge, Z.; Song, Z. Distributed PCA model for plant-wide process monitoring. *Ind. Eng. Chem. Res.* **2013**, *52*, 1947–1957. [CrossRef]
37. Bathelt, A.; Ricker, N.L.; Jelali, M. Revision of the Tennessee Eastman Process Model. *IFAC-PapersOnLine* **2015**, *48*, 309–314. [CrossRef]
38. Zeng, Z.; Gui, W.; Chen, X.; Xie, Y.; Zhang, H.; Sun, Y. A Cell Condition-Sensitive Frequency Segmentation Method Based on the Sub-Band Instantaneous Energy Spectrum of Aluminum Electrolysis Cell Voltage. *Engineering* **2021**, *7*, 1282–1292. [CrossRef]