Fault Detection Method Based on Global-Local Marginal Discriminant Preserving Projection for Chemical Process

Yang Li [1], Fangyuan Ma [1,2], Cheng Ji [1], Jingde Wang [1,*] and Wei Sun [1,*]

1   College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing 100029, China; 2019200013@mail.buct.edu.cn (Y.L.); 2020400053@mail.buct.edu.cn (F.M.); 2020400007@mail.buct.edu.cn (C.J.)

2   Center of Process Monitoring and Data Analysis, Wuxi Research Institute of Applied Technologies, Tsinghua University, Wuxi 214072, China

*   Correspondence: jingdewang@mail.buct.edu.cn (J.W.); sunwei@mail.buct.edu.cn (W.S.); Tel.: +86-010-6444-5826 (W.S.)

Abstract: Feature extraction plays a key role in fault detection methods. Most existing methods focus on comprehensive and accurate feature extraction of normal operation data to achieve better detection performance. However, discriminative features based on historical fault data are usually ignored. Aiming at this point, a global-local marginal discriminant preserving projection (GLMDPP) method is proposed for feature extraction. Considering its comprehensive consideration of global and local features, global-local preserving projection (GLPP) is used to extract the inherent feature of the data. Then, multiple marginal fisher analysis (MMFA) is introduced to extract the discriminative feature, which can better separate normal data from fault data. On the basis of fisher framework, GLPP and MMFA are integrated to extract inherent and discriminative features of the data simultaneously. Furthermore, fault detection methods based on GLMDPP are constructed and applied to the Tennessee Eastman (TE) process. Compared with the PCA and GLPP method, the effectiveness of the proposed method in fault detection is validated with the result of TE process.

Keywords: fault detection; discriminative feature extraction; multiple marginal fisher analysis; global local preserving projection

1. Introduction

The large scale and high complexity of modern chemical processes lead to greater challenges to the safety and stability of the processes. The fault detection and diagnose (FDD) methods, which can reduce the occurrence of faults and related losses, has received more attention [1]. As the first step in FDD and the basis for subsequent analysis, fault detection methods provide operators with real-time monitoring of process status and more response time for fault decision. With the widespread application of advanced measurement instruments and distribution control system (DCS), a large amount of data are collected and stored, which provides a good basis for the development of data-based fault detection methods [2].

In past decades, multivariate statistical process monitoring (MSPM) methods have been the most intensively and widely studied data-driven fault detection methods [3]. MSPM methods obtain the projection matrix to extract data feature by considering the statistic feature variability of data, such as variance for PCA-based methods and high-order statistic for ICA-based methods. Further, extended methods for the dynamic, nonlinearity and other characteristics of process data are proposed to extract more accurate features of data for better detection performance [4–6]. In fact, how to accurately extract the important features from massive data is the concern of most studies. However, the MSPM methods only focus on the global features of the data and the absence of local features represented by neighborhood information may compromise detection performance [7].

Fortunately, the emergence of manifold learning provides a novel perspective to preserve the local feature of data. Several methods based on manifold learning have been proposed for dimension reduction in pattern recognition, such as Laplacian eigenmap (LE), locally linear embedding (LLE) and local tangent space alignment (LTSA) [8–10]. Regarding the Out-of-Sample learning problem, their linear forms with explicit projection were proposed and have been introduced to the field of fault detection, such as local preserving projection (LPP), neighborhood preserving embedding (NPE), and linear local tangent space alignment (LLTSA) [11–13]. The above manifold learning-based methods extract local features represented by neighborhood information, while ignoring the global features expressed by variance information and high-order statistics.

In order to solve the issues of MSPM and manifold learning methods in features extraction, several feature extraction methods which can extract global and local features of data simultaneously were proposed and applied to fault detection. Zhang et al. first proposed the global-local structure analysis (GLSA) by integrating the objective functions of PCA and LPP directly [14]. Similarly, Yu proposed the local and global principal component analysis (LGPCA) by constructing objective function based on the ratio of LPP to PCA [15]. However, the PCA model in the above methods required data with Gaussian distribution. For this issue, Luo proposed a unified framework, namely, global-local preserving projection (GLPP) to extract global and local features based on the distance relationship between neighbors and non-neighbors entirely [16]. With different forms of local feature extraction, NPE and LTSA were later extended to extract global and local features of the data for fault detection [17–19]. Due to its lesser limitations on data distribution and low computational complexity, various GLPP-based improvements have been proposed, including dynamic, nonlinearity, non-parameterization, sparsity, and ensemble learning [20–27].

Most data-based fault detection methods, including MSPM-based and manifold learning-based methods, only rely on the more comprehensive and accurate feature extraction of normal operating conditions data to achieve better detection performance, which can contribute to detect any anomalies that exceed the defined normal range in the feature space obtained by normal data. However, data collected from normal operating conditions cannot be guaranteed to contain all features in normal operating conditions, which is a great challenge, and not always necessary either, as the purpose of fault detection is actually to distinguish the normal and abnormal operating condition based on data analysis. Due to the absence of discriminative features based on fault data, the feature extracted by the above methods may not be the optimal feature to distinguish normal operation conditions from real faults. Therefore, it is necessary to improve the performance of detection methods by considering fault data. Huang et al. proposed a novel slow feature analysis-based detection method and online fault information is used to reorder and select features online for obtaining fault-related feature [28]. However, the features selected by the above method are still derived from the slow features based on normal data. In terms of the above issue, the discriminative feature extraction method based on fault data should be introduced into the feature extraction of fault detection methods. Discriminant feature extraction methods represented by linear discriminant analysis (LDA) and its variants have been first applied in pattern recognition [29–31]. Due to the limitation of LDA-based methods on data distribution, marginal fisher analysis (MFA) based on graph embedding framework was proposed by maximizing the separability between pairwise marginal data points [32]. On this basis, multiple marginal fisher analysis (MMFA) was proposed to solve the class-isolation issue by considering the multiple marginal data pairs [33]. However, there is little discussion on how to combine discriminative feature methods with inherent feature extraction methods to improve performance of fault detection.

In this paper, a novel feature extraction algorithm, which is named global-local marginal discriminant preserving projection (GLMDPP), is proposed and applied for fault detection. Due to its ability to extract global and local features of the data simultaneously, inherent features of both normal data and historical fault data are extracted by GLPP method. Inspired by GLPP, discriminative features extraction based on marginal

sample pairs in MMFA is also extended to non-marginal sample pairs. Then, the objective functions of GLPP-based inherent features extraction and MMFA-based discriminative features extraction are integrated to obtain the optimal features which can separate normal conditions from historical fault while retaining full inherent features of the data. In addition, geodesic distance is introduced instead of Euclidean distance between non-neighbor or non-marginal sample pairs to represent intrinsic geometric structure more accurately. Statistics representing changes in feature space is calculated to establish a GLMDPP-based fault detection method.

The rest of paper is organized as follows. The basic methods related to the proposed method are briefly reviewed in Section 2. Section 3 presents the proposed GLMDPP method. GLMDPP-based fault detection procedure is developed in Section 4. The experimental results of the Tennessee Eastman process are discussed in Section 5. Section 6 provides the conclusion.

2. Preliminaries

2.1. Global-Local Preserving Projection

GLPP is a manifold learning-based feature extraction method which can preserving global and local features of data simultaneously [16]. In brief, GLPP extends the LPP-based adjacent relationship to the non-adjacent relationship to extract the comprehensive features of the data. The implementation procedures of GLPP are as follows:

Given a normalized data set $X = [x_1, x_2, \cdots, x_n] \in R^{m \times n}$, GLPP aims to obtain a projection matrix $A = \left[\vec{a}_1, \vec{a}_2, \cdots \vec{a}_d \right] \in R^{m \times d}$ which could map the X to $Y = A^T X = \left[\vec{y}_1, \vec{y}_2, \cdots \vec{y}_n \right] \in R^{d \times n}$. For each sample $\vec{x}_i$, the k nearest neighbors based on Euclidean distance are selected to construct local neighborhood $\Omega\left(\vec{x}_i \right) = [x_{i1}, x_{i2}, \cdots, x_{ik}] \in R^{m \times k}$. An adjacency weight matrix is constructed, and each element W_{ij} representing adjacency relationship between neighbor sample pairs is calculated by a heat kernel function as shown in Equation (1).

$$W_{ij} = \left\{ \begin{array}{ll} e^{\frac{-\|\vec{x}_i - \vec{x}_j\|^2}{\sigma_1}} & if\ x_j \in \Omega\left(\vec{x}_i \right)\ or\ x_i \in \Omega\left(\vec{x}_j \right) \\ 0 & otherwise \end{array} \right\} \tag{1}$$

where σ_1 is an empirical constant and $\|\ \|$ represents the Euclidean distance. LPP-based sub-objective function for local feature extraction is presented as follows:

$$\begin{aligned} J_{Local}\left(\vec{a} \right) &= \min_{\vec{a}} \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{ij} \\ &= \min_{\vec{a}} \frac{1}{2} \left\{ \sum_i y_i D_{ii} y_i^T - \sum_{ij} y_i W_{ij} y_j^T \right\} \\ &= \min_{\vec{a}} \frac{1}{2} \left\{ \sum_i \vec{a}^T \vec{x}_i D_{ii} \vec{x}_i^T \vec{a} - \sum_{ij} \vec{a}^T \vec{x}_i W_{ij} \vec{x}_i^T \vec{a} \right\} \\ &= \min_{\vec{a}} \frac{1}{2} \vec{a}^T X (D - W) X^T \vec{a} \\ &= \min_{\vec{a}} \frac{1}{2} \vec{a}^T X L X^T \vec{a} \end{aligned} \tag{2}$$

where D is a diagonal matrix and each element D_{ii} can be calculated by $D_{ii} = \sum\limits_{j} W_{ij}$. L represents Laplacian matrix which can be calculated by $L = D - W$. Similar to the local feature extraction part, non-adjacency weight matrix is constructed as follows:

$$\overline{W}_{ij} = \left\{ \begin{array}{ll} e^{\frac{-\|\vec{x}_i - \vec{x}_j\|^2}{\sigma_2}} & if \ x_j \notin \Omega\left(\vec{x}_i\right) \ and \ x_i \notin \Omega\left(\vec{x}_j\right) \\ 0 & otherwise \end{array} \right\} \tag{3}$$

where σ_2 and $\| \ \|$ in $\overline{W}_{ij}$ represent the same meaning with W_{ij}. On the basis of the non-adjacency relationship, the sub-objective function for global feature extraction is presented as follows:

$$\begin{aligned} J_{Global}\left(\vec{a}\right) &= \min_{\vec{a}} -\frac{1}{2}\sum_{ij}(y_i - y_j)^2 \overline{W}_{ij} \\ &= \min_{\vec{a}} - \left\{ \sum_i y_i \overline{D}_{ii} y_i^T - \sum_{ij} y_i \overline{W}_{ij} y_j^T \right\} \\ &= \min_{\vec{a}} - \left\{ \sum_i \vec{a}^T \vec{x}_i \overline{D}_{ii} \vec{x}_i^T \vec{a} - \sum_{ij} \vec{a}^T \vec{x}_i \overline{W}_{ij} \vec{x}_i^T \vec{a} \right\} \\ &= \min_{\vec{a}} - \vec{a}^T X\left(\overline{D} - \overline{W}\right) X^T \vec{a} \\ &= \min_{\vec{a}} - \vec{a}^T X \overline{L} X^T \vec{a} \end{aligned} \tag{4}$$

where $\overline{D}$ is a diagonal matrix and each element $\overline{D}_{ii}$ can be calculated by $\overline{D}_{ii} = \sum\limits_{j} \overline{W}_{ij}$. $\overline{L}$ represents Laplacian matrix which can be calculated by $\overline{L} = \overline{D} - \overline{W}$.

In order to preserve both global and local features of the data, a weighted coefficient η is introduced to integrate the two sub-objective functions as follows:

$$\begin{aligned} J_{GLPP}\left(\vec{a}\right) &= \min_{\vec{a}} \left\{ \eta J_{Local}\left(\vec{a}\right) - (1 - \eta) J_{Global}\left(\vec{a}\right) \right\} \\ &= \frac{1}{2} \left\{ \eta \sum_{ij}(y_i - y_j)^2 W_{ij} - (1 - \eta)\sum_{ij}(y_i - y_j)^2 \overline{W}_{ij} \right\} \\ &= \min_{\vec{a}} \frac{1}{2}\sum_{ij}(y_i - y_j)^2 R_{ij} \\ &= \min_{\vec{a}} \left\{ \sum_i y_i H_{ii} y_i^T - \sum_{ij} y_i R_{ij} y_j^T \right\} \\ &= \min_{\vec{a}} \left\{ \sum_i \vec{a}^T \vec{x}_i H_{ii} \vec{x}_i^T \vec{a} - \sum_{ij} \vec{a}^T \vec{x}_i R_{ij} \vec{x}_i^T \vec{a} \right\} \\ &= \min_{\vec{a}} \vec{a}^T X(H - R) X^T \vec{a} \\ &= \min_{\vec{a}} \vec{a}^T X M X^T \vec{a} \end{aligned} \tag{5}$$

where $R_{ij} = \eta W_{ij} - (1 - \eta)\overline{W}_{ij}$. H is a diagonal matrix and each element can be calculated by $H_{ii} = \sum\limits_{j} R_{ij}$. M is a Laplacian matrix which can be calculated by $M = H - R$. The weighted coefficient η is determined based on the trade-off between the local and global features as follows:

$$\eta = \frac{\rho(L)}{\rho(L) + \rho(\overline{L})} \tag{6}$$

where ρ denotes the spectral radius of the matrix. To avoid the singularity problem, the constraint is presented for the objective function of GLPP as follows:

$$\vec{a}^T \left(\eta X H X^T + (1 - \eta) I \right) \vec{a} = \vec{a}^T N \vec{a} = 1 \tag{7}$$

The optimization problem consisting of Equations (5) and (7) can be transformed to the generalized eigenvalue decomposition problem as follows:

$$X M X^T \vec{a} = \lambda N \vec{a} \tag{8}$$

Projection matrix A for preserving both global and local features of the data can be constructed by the eigenvectors corresponding to the d smallest eigenvalues.

2.2. Multiple Marginal Fisher Analysis

MMFA is a novel discriminative feature extraction method by maxing the interclass separability among multiple marginal point pairs and minimizing within-class scatter simultaneously [33]. Compared with LDA and MFA, limitations on Gaussian distribution of data and class-isolated issue are solved for better discriminative feature and wider applications.

Given a data set $X = [x_1, x_2, \cdots, x_n] \in R^{m \times n}$ and each sample x_i has its class label $l_i \in \{1, 2, \cdots c\}$. By using the projection matrix $A = \left[\vec{a}_1, \vec{a}_2, \cdots \vec{a}_d \right] \in R^{m \times d}$, the discriminative features can be obtained as follows:

$$Y = A^T X = \left[\vec{y}_1, \vec{y}_2, \cdots, \vec{y}_n \right] \in R^{d \times n} \tag{9}$$

For the within-class relationship, K-nearest neighbor method is used to determine the nearest neighbor relationship with the same class. The similarity between pairs of nearest neighbor points with the same class label is defined as follows:

$$C_{ij} = \left\{ \begin{array}{ll} \| \vec{x}_i - \vec{x}_j \|^2 & if\ l_i = l_j\ and\ \left\{ x_j \in \Omega \left(\vec{x}_i \right)\ or\ x_i \in \Omega \left(\vec{x}_j \right) \right\} \\ 0 & \text{otherwise} \end{array} \right\} \tag{10}$$

On the basis of above similarity, within-class compactness is characterized as follows:

$$\begin{aligned} S_w &= \min_{\vec{a}} \frac{1}{2} \sum_{ij} \| y_i - y_j \|^2 C_{ij} \\ &= \min_{\vec{a}} \vec{a}^T X (D_w - C) X^T \vec{a} \\ &= \min_{\vec{a}} \vec{a}^T X L_w X^T \vec{a} \end{aligned} \tag{11}$$

where D_w is a diagonal matrix and its element can be calculated by $D_{wii} = \sum_j C_{ij}$. The L_w is Laplacian matrix and it can be obtained by $L_w = D_w - C$. Then, distance-based k_2 nearest-neighbor sample pairs separability between every two classes are determined as follows:

$$\overline{C}_{ij} = \left\{ \begin{array}{ll} \| \vec{x}_i - \vec{x}_j \|^2 & if\ l_i \neq l_j\ and\ \| \vec{x}_i - \vec{x}_j \| \in k\ \text{nearest pairs between } l_i \text{ and } l_j \\ 0 & \text{otherwise} \end{array} \right\} \tag{12}$$

Furthermore, the interclass separability is defined by the above-mentioned nearest-neighbor sample pairs as follows:

$$
\begin{aligned}
S_b &= \max_{\vec{a}} \tfrac{1}{2} \sum_{ij} \| y_i - y_j \|^2 \overline{C}_{ij} \\
&= \max_{\vec{a}} \vec{a}^T X \left(D_b - \overline{C} \right) X^T \vec{a} \\
&= \max_{\vec{a}} \vec{a}^T X L_b X^T \vec{a}
\end{aligned}
\tag{13}
$$

On the basis of the interclass separability and within-class similarity defined in Equations (5) and (13), the following objective function is proposed by Fisher criterion.

$$
J_{MMFA} \left(\vec{a} \right) = \max_{\vec{a}} \frac{\vec{a}^T X L_b X^T \vec{a}}{\vec{a}^T X L_w X^T \vec{a}}
\tag{14}
$$

The above-mentioned objective function can be transformed into a solution of the generalized eigenvalue problem as follows:

$$
X L_b X^T \vec{a} = \lambda X L_w X^T \vec{a}
\tag{15}
$$

The optimal projection matrix A consists of the eigenvectors corresponding to the d largest eigenvalues.

3. GLMDPP Method

3.1. Inherent Feature Extraction

Due to its ability to comprehensively extract data features and the absence of limitation on data distribution, GLPP is applied for the inherent features extraction of data in the proposed algorithm. Given a dataset $X = [X_0, X_1, \cdots, X_c] \in R^{m \times n}$ and each class of dataset $X_i \in R^{m \times n_c}$. X_0 denotes the normal condition data and the rest data denotes the c class of historical fault data. The K-nearest neighbors method based on Euclidean distance is used to determine the nearest neighbor relationship within each class of data in historical data. Compared to Euclidean distance, geodesic distance is introduced due to its more accurate estimation of the non-neighborhood distance on the data manifold [34]. The Dijkstra algorithm was used to estimate the geodesic distance by calculating the shortest path distance based on the adjacency relationship [35]. The adjacency weight matrix of GLPP is shown in Equation (1) and the Euclidean distance of non-adjacency weight matrix is replaced with the geodesic distance in Equation (3). Additionally, the value of the σ_1 and σ_2 is determined by the average value of the Euclidean distance in adjacency relationship and geodesic distance in non-adjacency relationship, respectively. Then, the sub-objective functions of global and local feature extraction are presented by Equations (2) and (4). Due to the consideration of historical fault data, the adjacency weight matrix W and diagonal matrix D of local feature extraction sub-objective function can be written as:

$$
W = \begin{bmatrix} W_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_c \end{bmatrix}
\tag{16}
$$

$$
D = \begin{bmatrix} D_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & D_c \end{bmatrix}
\tag{17}
$$

where $W_0, \cdots, W_c$ represent the adjacency weight matrix in each class, and $D_0, \cdots, D_c$ represent a sample importance in its class. Non-adjacency weight matrix $\overline{W}$ and its corre-

sponding matrix D have the same form as above. Consistent with Equation (8), the objective function to extract both global and local features of the data is established as follows:

$$
\begin{aligned}
J_{inherent}\left(\overrightarrow{a}\right) &= \min_{\overrightarrow{a}}\left\{\eta J_{\text{Local}}\left(\overrightarrow{a}\right) - (1-\eta)J_{Global}\left(\overrightarrow{a}\right)\right\} \\
&= \min_{\overrightarrow{a}}\overrightarrow{a}^{T} XMX^{T}\overrightarrow{a}
\end{aligned}
\tag{18}
$$

3.2. Discriminative Feature Extraction

To improve the defined feature subspace by considering only normal conditions, fault conditions in the historical data are also taken into account to obtain the optimal discriminative feature subspace. Different with MMFA which considers marginal sample pairs for any two classes of data, only k_2 nearest neighbor sample pairs between normal data and each class of faulty conditions are considered as marginal sample pairs between classes. For subsequent integration, the weight form of the marginal sample pairs was adjusted to the form of a heat kernel function consistent with GLPP as Equation (19). Inspired from GLPP, non-marginal sample pairs are also introduced to extract the discriminative features comprehensively. Since marginal sample pairs have a more important role for partial overlap inter-class data in the original space, the form of weights for non-marginal sample pairs is constructed as follows:

$$
C_{ij} = \begin{cases}
e^{\frac{-\|\overrightarrow{x}_i - \overrightarrow{x}_j\|^2}{\sigma_3}} & if\ l\left(\overrightarrow{x}_i\right) \neq l\left(\overrightarrow{x}_j\right) \text{and } \|\overrightarrow{x}_i - \overrightarrow{x}_j\| \in \min \text{ k nearest sample pairs} \\
e^{\frac{-D_G(x_i,x_j)^2}{\sigma_4}}\left(1 - e^{\frac{-D_G(x_i,x_j)^2}{\sigma_4}}\right) & if\ l\left(\overrightarrow{x}_i\right) \neq l\left(\overrightarrow{x}_j\right) \text{and } \|\overrightarrow{x}_i - \overrightarrow{x}_j\| \notin \min \text{ k nearest sample pairs}
\end{cases}
\tag{19}
$$

where $\|\ \|$ and $D_G\left(x_i, x_j\right)$ represents the Euclidean distance and geodesic distance between x_i and x_j, respectively. Additionally, the value of the σ_3 and σ_4 is identified by the average value of the Euclidean distance in marginal sample pairs and geodesic distance in non-marginal sample pairs, respectively. Furthermore, sub-objective function for discriminative feature extraction is presented as follows:

$$
\begin{aligned}
J_{discriminal}\left(\overrightarrow{a}\right) &= \min_{\overrightarrow{a}}\frac{1}{2}\sum_{ij}\left(y_i - y_j\right)^2 C_{ij} \\
&= \min_{\overrightarrow{a}}\left\{\sum_i y_i D_{w_{ii}} y_i^{T} - \sum_{ij} y_i C_{ij} y_j^{T}\right\} \\
&= \min_{\overrightarrow{a}}\left\{\sum_i \overrightarrow{a}^{T}\overrightarrow{x}_i D_{w_{ii}}\overrightarrow{x}_i^{T}\overrightarrow{a} - \sum_{ij}\overrightarrow{a}^{T}\overrightarrow{x}_i C_{ij}\overrightarrow{x}_i^{T}\overrightarrow{a}\right\} \\
&= \min_{\overrightarrow{a}}\overrightarrow{a}^{T} X(D_w - C)X^{T}\overrightarrow{a} \\
&= \min_{\overrightarrow{a}}\overrightarrow{a}^{T} XLX^{T}\overrightarrow{a}
\end{aligned}
\tag{20}
$$

where D_w is a diagonal matrix and its element can be calculated by $D_{wii} = \sum_j C_{ij}$. The L is a Laplacian matrix which is obtained by $L = D_w - C$.

3.3. Formulation of FDGLPP

In order to preserve both the inherent and discriminant features of the data, two sub-objective functions are integrated by fisher criterion as follows:

$$
J_{GLMDPP}\left(\overrightarrow{a}\right) = \min\frac{J_{inherent}\left(\overrightarrow{a}\right)}{J_{discriminal}\left(\overrightarrow{a}\right)} = \min\frac{\overrightarrow{a}^{T} XMX^{T}\overrightarrow{a}}{\overrightarrow{a}^{T} XLX^{T}\overrightarrow{a}}
\tag{21}
$$

Then, the above optimization problem is transformed to a generalized eigenvalue problem by the Lagrange multiplier method as follows:

$$XMX^T\vec{a} = \lambda XLX^T\vec{a} \tag{22}$$

Optimal discriminative feature projection matrix A is obtained by the eigenvectors corresponding to the d largest eigenvalues.

4. GLMDPP-Based Fault Detection

Given the dataset $X = [X_0, X_1, \cdots, X_c] \in R^{m \times n}$, Z-Score standardization is used to normalize the dataset X with the mean and standard deviation of normal dataset $X_0 \in R^{m \times n_0}$. Discriminative feature projection matrix A is obtained by GLMDPP method and discriminative feature of the new sample x_{new} can be calculated as follows:

$$y_{new} = A^T X_{new} \in R^{d \times 1} \tag{23}$$

Since the discriminative feature based on historical normal and fault data is already the most sensitive feature space for detecting faults, only the T^2 statistic is introduced for monitoring the variation of feature space as follows:

$$T_{new}^2 = \vec{y}_{new}^T S^{-1} \vec{y}_{new} \tag{24}$$

where $S = Y^T Y/(n-1)$ is the covariance matrix of the discriminative feature Y projected from the normal data X_0. The T^2 statistic control limit is presented as follows:

$$T_\alpha^2 = \frac{d(n-1)}{n-d} F_\alpha(d, n-d) \tag{25}$$

where α represents the significance level and $F(d, n-d)$ denotes a F distribution with d and n-d degrees of freedom. If the T^2 statistic of a new sample exceeds its control limit, the process is considered as fault state; Otherwise, it is considered as normal state.

The complete procedure of GLMDPP-based fault detection method is shown in Figure 1, including two parts: offline modeling and online detection.
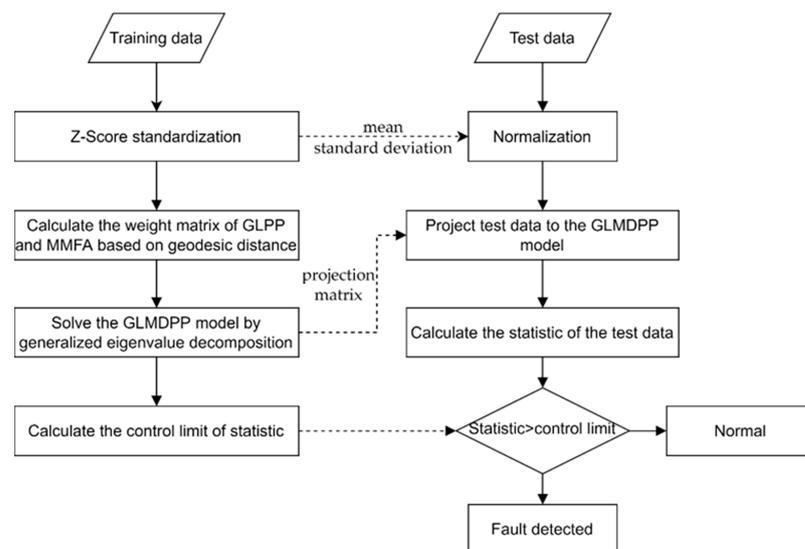

Figure 1. Fault detection procedure based on GLMDPP algorithm.

Offline modelling:

(1) Historical data including fault data and normal data is used as training data and Z-Score standardization is employed for normalize the training data via the mean $X_{0,mean}$ and standard deviation $X_{0,std}$ of normal data as follows:

$$X' = \frac{X - X_{0,mean}}{X_{0,std}} \tag{26}$$

(2) The Euclidean distance-based adjacency weight matrix and marginal sample pairs weight matrix are constructed. Based on the adjacency relationship, the geodesic distance is introduced to construct the non-adjacency weight matrix and the non-marginal sample pairs weight matrix.

(3) On the basis of GLPP and MMFA, the objective function of GLMDPP which can extract both inherent and discriminant features simultaneously is constructed by fisher criterion.

(4) The objective function of GLMDPP is solved by transforming to a generalized eigenvalue problem and projection matrix A is obtained.

(5) The control limit of T^2 statistics is calculated as shown in Equation (25)

Online monitoring:

(1) Online test data is collected and normalized with the mean and variance of the normal training data.

(2) The feature of test data is calculated by the projection matrix A obtained from offline modelling.

(3) T^2 statistics is calculated and compared with the control limit.

(4) If T^2 statistic of online test data exceeds its control limit, fault is detected. Otherwise, return to (1).

5. Case Study

In order to demonstrate the fault detection performance of the proposed method, the TE process is used as a benchmark test [36]. The flow diagram of the TE process is shown in Figure 2. It contains five major unit operations, which are the reactor, the product condenser, a vapor–liquid separator, a recycle compressor and a product strip. The TE process involves 52 variables, including 41 measured variables and 11 manipulated variables. Due to 19 quality measured variables being sampled less frequently, the remaining 22 measured variables and 11 manipulated variables sampled every 3 min are usually utilized as monitoring variables. In the TE process, 21 types of faults can be generated to test the detection performance and the description of each fault can be seen in Table 1. The variable information of TE process is listed in Table 2.
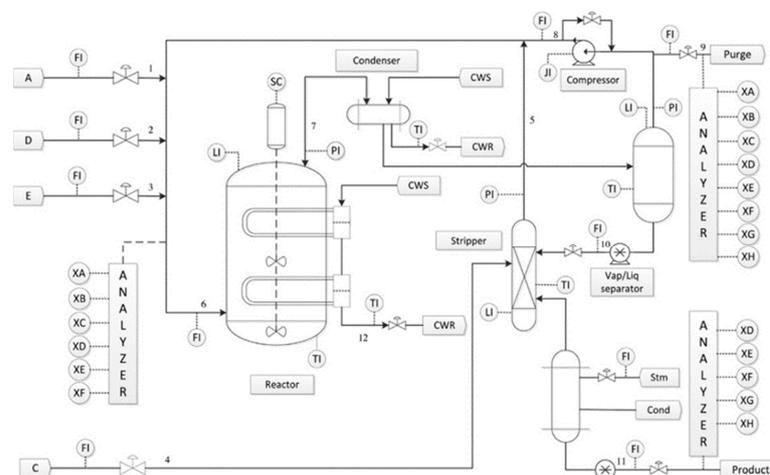

Figure 2. Flow diagram of the TE process.

Table 1. Programmed faults in the TE process.

No.	Fault Description	Type
1	A/C feed ratio, B composition constant (stream 4)	step
2	B composition, A/C ratio constant (stream 4)	step
3	D feed temperature (stream 2)	step
4	reactor cooling water inlet temperature	step
5	condenser cooling water inlet temperature	step
6	A feed loss (stream 1)	step
7	C header pressure loss-reduced availability (stream 4)	step
8	A, B, C feed composition (stream 4)	random variation
9	D feed temperature (stream 2)	random variation
10	C feed temperature (stream 4)	random variation
11	reactor cooling water inlet temperature	random variation
12	condenser cooling water inlet temperature	random variation
13	reaction kinetics	slow drift
14	reactor cooling water valve	sticking
15	condenser cooling water valve	sticking
16	unknown	unknown
17	unknown	unknown
18	unknown	unknown
19	unknown	unknown
20	unknown	unknown
21	the valve for stream 4	constant position

Table 2. Variable information of TEP for process monitoring.

Variable	Description	Variable	Description
F1	A feed (stream 1)	T18	Stripper temperature
F2	D feed (stream 2)	F19	Stripper steam flow
F3	E feed (stream 3)	C20	Compressor work
F4	A and C feed (stream 4)	T21	Reactor cooling water outlet temperature
F5	Recycle flow (stream 8)	T22	Separator cooling water outlet temperature
F6	Reactor feed rate (stream 6)	V23	D feed flow (stream 2)
P7	Reactor pressure	V24	E feed flow (stream 3)
L8	Reactor level	V25	A feed flow (stream 1)
T9	Reactor temperature	V26	A and C feed flow (stream 4)
F10	Purge rate (stream 9)	V27	Compressor recycle valve
T11	Product separator temperature	V28	Purge valve (stream 9)
L12	Product separator level	V29	Separator pot liquid flow (stream 10)
P13	Product separator pressure	V30	Stripper liquid prod flow (stream 11)
F14	Product separator underflow (stream 10)	V31	Stripper steam valve
L15	Stripper level	V32	Reactor cooling water flow
P16	Stripper pressure	V33	Condenser cooling water flow
F17	Stripper underflow (stream 11)		

The data applied in this research is provided in Prof. Braatz's homepage in MIT and it can be divided into training data and test data [37]. Normal data with 500 samples and 21 types of fault data with 480 samples each are used as training data. Corresponding with the above faults, 21 test datasets with 960 samples are used as test data and fault is introduced into the 161st sample in each dataset.

In order to illustrated the detection performance of the proposed method, GLPP-based and PCA-based fault detection methods is introduced for comparison. Similar to PCA, cumulative percent variance of 90% is used to determine the number (d) of projection vectors for three methods. The same number of nearest neighbors $k_1 = 10$ is chosen for both GLPP and GLMDPP. The number of marginal sample pairs $k_2 = 50$ is determined empirically for GLMDPP. The confidence level for control limit of T^2 statistic is set as 99% for three methods. To compare the detection performance of different methods quantitatively,

fault detection rate (FDR) and false alarm rate (FAR) are adopted and their calculation is presented as follows:

$$FDR = \frac{x \in fault : T^2(x) > limit}{normal\ samples} \times 100\% \tag{27}$$

$$FAR = \frac{x \in normal : T^2(x) > limit}{normal\ samples} \times 100\% \tag{28}$$

The fault detection rates (FDRs) and false alarm rates (FARs) of three monitoring methods for TE process are shown in Table 3. Previous studies have shown that faults 3, 9 and 15 are difficult to detect due to their small magnitudes [37]. The detection performance of the rest 18 faults is compared and their average values are calculated. The bolded numbers in Table 3 represent the best FDR for the corresponding statistic under each fault. Obviously, the proposed method gives the highest detection rate for 17 of 18 faults, especially for fault 10, 16 and 21 with significant improvement. Compared with PCA that considers only global features of the data, the proposed method has higher detection rates on almost all faults. Compared to GLPP, the proposed method has the highest detection rate for 17 faults except for fault 2, and there is little difference on detection rate of fault 2. The performance of the proposed method can be further demonstrated by the average FDR of 18 faults given in Table 3. False alarm rate (FAR) is also an important indicator for evaluating the detection performance. It is easy to see that the of the proposed method has a low average FAR, which is only slightly higher than GLPP. In addition, fault detection time of different methods are presented in Table 4 and the earliest valid fault detection time is highlighted in bold. Compared with PCA, GLPP and the proposed method have outstanding performance in fault detection time. Compared with GLPP, the proposed method has the same detection time on most faults and earlier detection time especially on faults 8 and 12. The difference between the proposed method and the comparative methods in fault detection performance is due to the introduction of known fault data in the modeling process. Specifically, by combining inherent feature extraction method with discriminant feature extraction method, historical data including normal condition data and known faults data are merged into the training data, so as to achieve better detection performance on faults with similar characteristics to known cases. In general, the proposed method has superior detection performance among three methods due to its comprehensive extraction of inherent feature and discriminative feature.

To further illustrate the effectiveness of the proposed method, fault detection results of three methods in fault 21 are presented in Figure 3. It can be found that the PCA method detects the fault at 627th sample and has the lowest FDR. Both GLPP and GLMDPP can detect this fault at 417th sample simultaneously. Compared with GLPP, GLMDPP not only has a higher FDR but also a lower FAR after detecting the fault.

**Table 3.** FDRs (%) and FARs (%) of different method on TE process.

No.	PCA		GLPP		GLMDPP	
	FDR	**FAR**	**FDR**	**FAR**	**FDR**	**FAR**
1	99.25	0.63	**100.00**	0.63	**100.00**	0.63
2	98.25	1.25	**99.13**	0.63	99.00	0.63
3	5.75	1.25	6.00	3.13	7.25	3.13
4	68.13	1.25	**100.00**	0.63	**100.00**	1.25
5	27.75	1.25	**100.00**	0.63	**100.00**	1.25
6	99.50	0.63	**100.00**	0.00	**100.00**	0.00
7	**100.00**	1.88	**100.00**	2.50	**100.00**	1.88
8	97.25	0.63	98.25	0.00	**98.50**	0.63
9	5.63	10.00	3.88	6.88	6.13	14.38
10	44.50	2.50	87.25	1.88	**89.38**	1.88
11	60.75	1.88	80.38	1.25	**81.88**	1.25
12	98.50	1.25	99.75	1.88	**99.88**	1.25
13	94.38	0.00	95.38	0.00	**95.50**	0.00
14	**100.00**	1.25	**100.00**	0.63	**100.00**	1.88
15	7.75	0.00	12.63	0.63	17.38	1.25
16	29.75	12.50	91.38	5.63	**93.50**	8.75
17	84.75	1.25	96.75	1.88	**97.00**	3.13
18	89.63	1.88	90.50	1.88	**90.63**	1.88
19	15.88	0.00	92.88	0.00	**93.88**	0.00
20	43.00	0.63	88.00	0.00	**89.88**	0.00
21	43.50	1.88	55.38	4.38	**60.38**	3.75
average	71.93	1.81	93.06	1.35	**93.85**	1.67

Table 4. Fault detection time of different method on TE process.

No.	PCA	GLPP	GLMDPP
1	167	**161**	**161**
2	175	**171**	**171**
3	-	-	-
4	163	**161**	**161**
5	**161**	**161**	**161**
6	165	**161**	**161**
7	**161**	**161**	**161**
8	186	176	**174**
9	-	-	-
10	126	**182**	**182**
11	**166**	**166**	**166**
12	163	163	**162**
13	207	**201**	**201**
14	**161**	**161**	**161**
15	-	-	-
16	196	**167**	**167**
17	187	**182**	**182**
18	248	**178**	**178**
19	237	**170**	**170**
20	244	**223**	**223**
21	627	**417**	**417**
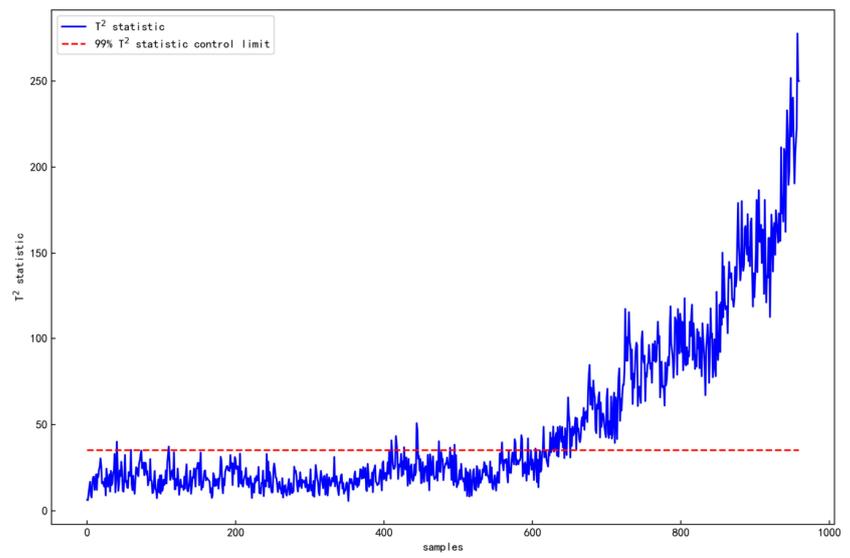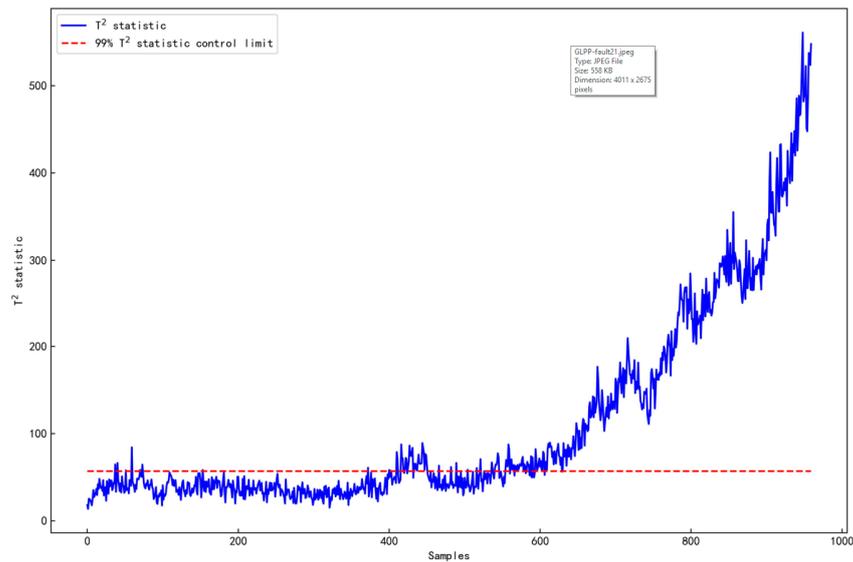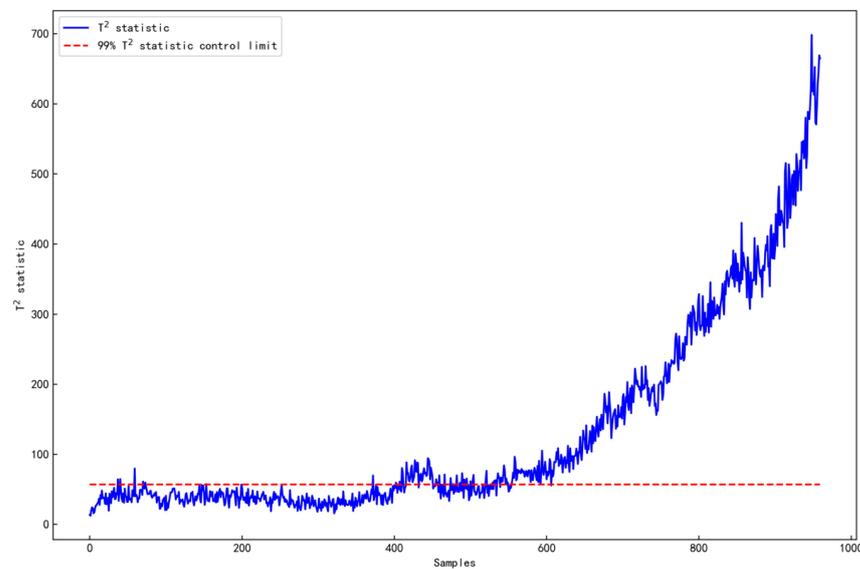
Figure 3. Detection charts of three methods for fault21: (**a**) PCA; (**b**) GLPP; (**c**) GLMDPP.

6. Conclusions

Regarding the absence of discriminative information in most fault detection methods based on normal condition data, a global-local marginal discriminant preserving projection (GLMDPP)-based fault detection method is proposed. Historical data, including normal and fault data, are used for the construction of the GLMDPP method. Specifically, GLPP and MMFA-based marginal relationships are integrated through the Fisher framework to extract both the inherent feature and discriminative feature of data, which is expected to have better performance on faults with similar characteristics to known cases. Comparing PCA-based and GLPP-based monitoring methods, the proposed monitoring method is tested by the TE process. The results confirm that the consideration of historical fault data can contribute to the improvement of fault detection performance which provides a new idea for improvement of fault detection method only based on normal data. Furthermore, the imbalance issue between historical fault data and normal data and multimodal characteristic in real processes will be studied in the future.

Author Contributions: Conceptualization, Y.L.; Methodology, Y.L.; Software, Y.L. and F.M.; Validation, Y.L., C.J. and W.S.; Formal Analysis, Y.L.; Investigation, Y.L., F.M. and C.J.; Resources, W.S. and J.W.; Data Curation, W.S. and Y.L.; Writing—Original Draft Preparation, Y.L.; Writing—Review and Editing, W.S., J.W., F.M. and Y.L.; Visualization, Y.L.; Supervision, W.S. and J.W.; Project Administration, W.S. and J.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data applied in this research can be obtained at https://github.com/camaramm/tennessee-eastman-profBraatz (accessed on 22 November 2021).

References

1. Qin, S.J.; Chiang, L.H. Advances and opportunities in machine learning for process data analytics. *Comput. Chem. Eng.* **2019**, *126*, 465–473. [CrossRef]
2. Ge, Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 16–25. [CrossRef]
3. Joe Qin, S. Statistical process monitoring: Basics and beyond. *J. Chemom. J. Chemom. Soc.* **2003**, *17*, 480–502. [CrossRef]
4. Ku, W.; Storer, R.H.; Georgakis, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179–196. [CrossRef]
5. Choi, S.W.; Lee, C.; Lee, J.-M.; Park, J.H.; Lee, I.-B. Fault detection and identification of nonlinear processes based on kernel PCA. *Chemom. Intell. Lab. Syst.* **2005**, *75*, 55–67. [CrossRef]
6. Lee, J.M.; Qin, S.J.; Lee, I.B. Fault detection and diagnosis based on modified independent component analysis. *AIChE J.* **2006**, *52*, 3501–3514. [CrossRef]
7. Hu, K.; Yuan, J. Multivariate statistical process control based on multiway locality preserving projections. *J. Process Control* **2008**, *18*, 797–807. [CrossRef]
8. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]
9. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, Vancouver, BC, Canada, 8–14 December 2001.
10. Zhang, Z.; Zha, H. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. *SIAM J. Sci. Comput.* **2004**, *26*, 313–338. [CrossRef]
11. He, X.; Niyogi, P. Locality preserving projections. *Adv. Neural Inf. Processing Syst.* **2003**, *16*, 153–160. [CrossRef]
12. He, X.; Cai, D.; Yan, S.; Zhang, H.-J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; Volume 1, pp. 1208–1213.
13. Zhang, T.; Yang, J.; Zhao, D.; Ge, X. Linear local tangent space alignment and application to face recognition. *Neurocomputing* **2007**, *70*, 1547–1553. [CrossRef]
14. Zhang, M.; Ge, Z.; Song, Z.; Fu, R. Global–Local Structure Analysis Model and Its Application for Fault Detection and Identification. *Ind. Eng. Chem. Res.* **2011**, *50*, 6837–6848. [CrossRef]
15. Yu, J. Local and global principal component analysis for process monitoring. *J. Process Control* **2012**, *22*, 1358–1373. [CrossRef]
16. Luo, L. Process Monitoring with Global–Local Preserving Projections. *Ind. Eng. Chem. Res.* **2014**, *53*, 7696–7705. [CrossRef]
17. Ma, Y.; Song, B.; Shi, H.; Yang, Y. Fault detection via local and nonlocal embedding. *Chem. Eng. Res. Des.* **2015**, *94*, 538–548. [CrossRef]

18. Dong, J.; Zhang, C.; Peng, K. A novel industrial process monitoring method based on improved local tangent space alignment algorithm. *Neurocomputing* **2020**, *405*, 114–125. [CrossRef]

19. Fu, Y. Local coordinates and global structure preservation for fault detection and diagnosis. *Meas. Sci. Technol.* **2021**, *32*, 115111. [CrossRef]

20. Luo, L.; Bao, S.; Mao, J.; Tang, D. Nonlinear process monitoring using data-dependent kernel global–local preserving projections. *Ind. Eng. Chem. Res.* **2015**, *54*, 11126–11138. [CrossRef]

21. Bao, S.; Luo, L.; Mao, J.; Tang, D. Improved fault detection and diagnosis using sparse global-local preserving projections. *J. Process Control* **2016**, *47*, 121–135. [CrossRef]

22. Luo, L.; Bao, S.; Mao, J.; Tang, D. Nonlocal and local structure preserving projection and its application to fault detection. *Chemom. Intell. Lab. Syst.* **2016**, *157*, 177–188. [CrossRef]

23. Zhan, C.; Li, S.; Yang, Y. Enhanced Fault Detection Based on Ensemble Global–Local Preserving Projections with Quantitative Global–Local Structure Analysis. *Ind. Eng. Chem. Res.* **2017**, *56*, 10743–10755. [CrossRef]

24. Tang, Q.; Liu, Y.; Chai, Y.; Huang, C.; Liu, B. Dynamic process monitoring based on canonical global and local preserving projection analysis. *J. Process Control* **2021**, *106*, 221–232. [CrossRef]

25. Huang, C.; Chai, Y.; Liu, B.; Tang, Q.; Qi, F. Industrial process fault detection based on KGLPP model with Cam weighted distance. *J. Process Control* **2021**, *106*, 110–121. [CrossRef]

26. Cui, P.; Wang, X.; Yang, Y. Nonparametric manifold learning approach for improved process monitoring. *Can. J. Chem. Eng.* **2022**, *100*, 67–89. [CrossRef]

27. Yang, F.; Cui, Y.; Wu, F.; Zhang, R. Fault Monitoring of Chemical Process Based on Sliding Window Wavelet DenoisingGLPP. *Processes* **2021**, *9*, 86. [CrossRef]

28. Huang, J.; Ersoy, O.K.; Yan, X. Slow feature analysis based on online feature reordering and feature selection for dynamic chemical process monitoring. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 1–11. [CrossRef]

29. Fisher, R.A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **1936**, *7*, 179–188. [CrossRef]

30. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [CrossRef]

31. Pylkkönen, J. LDA based feature estimation methods for LVCSR. In Proceedings of the Ninth International Conference on Spoken Language Processing, Pittsburgh, Pennsylvania, 17–21 September 2006.

32. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.-J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *29*, 40–51. [CrossRef]

33. Huang, Z.; Zhu, H.; Zhou, J.T.; Peng, X. Multiple Marginal Fisher Analysis. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9798–9807. [CrossRef]

34. Fu, Y.; Luo, C. Joint Structure Preserving Embedding Model and Its Application for Process Monitoring. *Ind. Eng. Chem. Res.* **2019**, *58*, 20667–20679. [CrossRef]

35. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [CrossRef]

36. Downs, J.J.; Vogel, E.F. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, *17*, 245–255. [CrossRef]

37. Chiang, L.; Russell, E.; Braatz, R. *Fault Detection and Diagnosis in Industrial Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2001.