Research Article

An Efficient Quality-Related Fault Diagnosis Method for Real-Time Multimode Industrial Process

Kaixiang Peng, Bingzheng Wang, and Jie Dong

Key Laboratory for Advanced Control of Iron and Steel Process, Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Jie Dong; dongjie@ies.ustb.edu.cn

Focusing on quality-related complex industrial process performance monitoring, a novel multimode process monitoring method is proposed in this paper. Firstly, principal component space clustering is implemented under the guidance of quality variables. Through extraction of model tags, clustering information of original training data can be acquired. Secondly, according to multimode characteristics of process data, the monitoring model integrated Gaussian mixture model with total projection to latent structures is effective after building the covariance description form. The multimode total projection to latent structures (MTPLS) model is the foundation of problem solving about quality-related monitoring for multimode processes. Then, a comprehensive statistics index is defined which is based on the posterior probability of the monitored samples belonging to each Gaussian component in the Bayesian theory. After that, a combined index is constructed for process monitoring. Finally, motivated by the application of traditional contribution plot in fault diagnosis, a gradient contribution rate is applied for analyzing the variation of variable contribution rate along samples. Our method can ensure the implementation of online fault monitoring and diagnosis for multimode processes. Performances of the whole proposed scheme are verified in a real industrial, hot strip mill process (HSMP) compared with some existing methods.

1. Introduction

With modern industrial processes getting increasingly complex and large, prevention monitoring and fault diagnosis have become the key to ensure safe operation, improve product quality, and gain economic benefits. Due to the complex operation mechanism, sheer size, complex conditions, chaotic environment, and vague boundary conditions in complex industrial systems, it is quite tough to implement effective process monitoring. As a result, the data-driven process monitoring technology has become one of the research hotspots in the field of fault diagnosis. The core idea of this technique is to establish the data model by means of using historical data, mining useful information, and getting the features of normal and fault operation mode, so as to realize process monitoring. In the last decades, basic multivariate statistical monitoring techniques, such as principal component analysis (PCA) and partial least squares (PLS), have been established and successfully applied in practice [1].

However, PCA or PLS model is established with data which follow the basis hypothesis of data subject to stable single Gaussian mode. Due to the reasons of fluctuation of raw materials, product specifications, and differences among batches, process data show the characteristic of multimode in actual industrial processes especially for batch processes. Considering the problems existing in the multimode process, traditional fault detection methods and their improved algorithms are difficult to be applied directly; otherwise, the performance of data model in process monitoring will be reduced.

Many scholars have studied a lot and made some progress on those problems [1]. Hwang and Han proposed a hierarchical clustering based on the PCA modeling method [2]. Lane et al. proposed a pooled principal component analysis method [3]. However, the ensemble modeling methods, in which the common feature of subspace in each mode is extracted as a unified model, are unable to fully or accurately depict all operation models. Particularly, when there are many

differences among various modes, the model characterization in their methods is often biased. Chen and Liu used the heuristic smoothing clustering algorithm to classify data automatically, which can get multiple operating modes [4]. Zhao et al. applied multiple PCA and multiple PLS method to fault monitoring for multimode processes [5], in which the similarity index between different operating models was established and used to analyze the shift between the models. In view of stage division, Doan and Srinivasan modeled different stages of the process, respectively, for fault monitoring [6]. Dealing with the multimode problem of the process, the former divided the process data using the clustering method and then established independent models, so as to make fault monitoring more targeted. However, the above independent modeling methods are often complex, have large calculating quantity, and are usually based on the experience of mode division. Whether the division is reasonable or not will directly affect the quality of monitoring results. All the above increase the difficulties of online monitoring.

Considering the unique advantages in dealing with non-Gaussian data, the Gaussian mixture model (GMM) has not been explored in multimode process monitoring until recently. Choi et al. integrated PCA and DA with GMM to detect and isolate the faults in a process with nonlinearity, multistates, or dynamics [7]. Yoo et al. applied a similar strategy into multiway PCA to monitor biological batch processes [8]. However, these methods ignore the possibility that the monitored sample may come from other Gaussian components of lower posterior probabilities, which may lead to biased monitoring results. Yu and Qin proposed a new method that combines finite mixture Gaussian models with Bayesian inference to characterize different operation modes through Gaussian components and then realized fault detection [9]. In recent years, many scholars had proposed different methods to solve multimode monitoring [9].

The main contribution of this paper is summarized as follows. (1) An efficient method for multimode process monitoring based on finite Gaussian mixture models is proposed. (2) A gradient contribution rate is proposed to measure the contribution to the combined index and find out the variable which should be in charge of the fault in quality. This rate can better show the changes of variables contribution rate over time after fault occurrence.

The remainder of this paper is organized as follows. In Section 2, the descriptions of traditional PCA and PLS models in covariance form are provided, and then the covariance description form of the total projection to latent structures (TPLS) model is derived. Multimode information is extracted from the principal component space by GMM and a new multimode total projection to potential structure (MTPLS) model is established in Section 3. A unified monitoring framework based on MTPLS in combination with Bayesian inference is constructed and quality-related fault monitoring is implemented using a combined index in Section 4. In Section 5, a hot strip mill process is taken as an example to verify the superiority of our new method in fault monitoring and diagnosis over traditional methods. The conclusions and future works are given in Section 6.

2. Multivariate Statistical Theory

2.1. PCA and Covariance Description Form. Principal component analysis model is one of the most basic projection models in multivariate statistical analysis. Let $\mathbf{X} \in \mathbf{R}^{N*m}$ be the dataset of m-dimensional process variables, where N stands for the number of samples. Matrix $\mathbf{X}$ can be decomposed into a score matrix and a loading matrix as follows [10]:

$$\mathbf{X} = \widehat{\mathbf{X}} + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E},$$
$$\mathbf{T} = \mathbf{X}\mathbf{P}, \tag{1}$$

where $\mathbf{T} \in \mathbf{R}^{N*A}$ and $\mathbf{P} \in \mathbf{R}^{m*A}$ stand for score matrix and loading matrix, respectively, and A is the number of principal components. The covariance matrix of normalized data can be defined as follows:

$$\Sigma_{\mathbf{X}} \approx \frac{1}{N-1}\mathbf{X}^T\mathbf{X}. \tag{2}$$

The PCA loading matrix $\mathbf{P}$ can be obtained by eigenvalue decomposition on the covariance matrix $\Sigma_{\mathbf{X}}$.

Based on the projection model, monitoring statistics indexes T^2 and SPE can be constructed. Let $\mathbf{x}_{\text{new}} \in R^m$; the indexes can be designed as follows:

$$T^2 = \mathbf{x}_{\text{new}}^T \mathbf{P}\Lambda^{-1}\mathbf{P}^T \mathbf{x}_{\text{new}} \leq T_\alpha^2,$$
$$\text{SPE} = \left\| \left(\mathbf{I} - \mathbf{P}\mathbf{P}^T\right)\mathbf{x}_{\text{new}} \right\|^2 \leq \delta_\alpha^2, \tag{3}$$

where Λ denotes the principal component covariance matrix and T_α^2 and δ_α^2 are the control limit with the confidence level of α.

When the residual error is subject to normal distribution, Jackson and Mudholkar pointed out that the control limit can be calculated as follows:

$$\delta_\alpha^2 = \theta_1 \left(\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{1/h_0}, \tag{4}$$

where $h_0 = 1 - 2\theta_1\theta_3/3\theta_1^2$, $\theta_i = \sum_{j=A+1}^{m} \lambda_j^i$ $(i = 1, 2, 3)$, c_α represents the threshold of standard normal distribution under the confidence level of α, and λ_j represents the eigenvalue of covariance matrix $\Sigma_{\mathbf{X}}$.

Similarly, in order to apply the sample covariance information into the monitoring index, the principal component covariance matrix can be expressed as

$$\Lambda = \frac{1}{N-1}\mathbf{T}^T\mathbf{T} = \frac{1}{N-1}\mathbf{P}^T\mathbf{X}^T\mathbf{X}\mathbf{P} = \mathbf{P}^T\Sigma_{\mathbf{X}}\mathbf{P}. \tag{5}$$

2.2. PLS and Covariance Description Form. In the actual industrial production, the changes of quality variables $\mathbf{Y}$ are of more concern, especially for the faults which can cause the change of quality variables. PLS model uses the quality variables to guide the decomposition of sample space.

PLS decomposition of $\mathbf{X}$ and $\mathbf{Y}$ results in the following:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E},$$
$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{F}, \tag{6}$$

where $\mathbf{X} \in \mathbf{R}^{N*m}$, $\mathbf{Y} \in \mathbf{R}^{N*l}$, and score matrix $\mathbf{T}$ can be formulated with $\mathbf{X}$ as $\mathbf{T} = \mathbf{XR}$.

Parameter matrix $\mathbf{R}$ can be obtained by the loading matrix $\mathbf{P}$ and weight matrix $\mathbf{W}$ in PLS iterative calculation, $\mathbf{R} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}$.

According to the iterative process of the PLS model, Peng et al. proposed a model construction method using data covariance information [11], in which the covariance matrix of data was introduced into the iterative process, and model parameter matrices can be obtained at the same time. Compared with conventional PLS, the model construction method using data covariance information reduced the calculation amount although the intrinsic properties essence was not changed.

Different from the PCA projection model, the decomposition structure of space $\mathbf{X}$ in PLS is defined by two matrices $\mathbf{P}$ and $\mathbf{R}$, and an oblique projection structure is induced in input space. It is the quality that guides the decomposition of sample space, so that the principal component space is changed. The covariance matrix of the principal component space can be expressed as

$$\mathbf{\Lambda} = \frac{1}{N-1}\mathbf{T}^T\mathbf{T} = \frac{1}{N-1}\mathbf{R}^T\mathbf{X}^T\mathbf{X}\mathbf{R} = \mathbf{R}^T\mathbf{\Sigma_X}\mathbf{R}. \tag{7}$$

Similar to PCA model monitoring, the monitoring sample statistics can be constructed by using the covariance matrix of the above formula as follows:

$$T^2 = \mathbf{x}_{\text{new}}^T\mathbf{R}\mathbf{\Lambda}^{-1}\mathbf{R}^T\mathbf{x}_{\text{new}} \leq T_\alpha^2,$$
$$Q = \left\|\left(\mathbf{I} - \mathbf{PR}^T\right)\mathbf{x}_{\text{new}}\right\|^2 \leq \delta_\alpha^2. \tag{8}$$

The control limit of the residual statistic can be calculated as follows:

$$\delta_\alpha^2 = g\chi_{h,\alpha}^2, \tag{9}$$

where $g = S/2\mu$, $h = 2\mu^2/S$, μ represents the sample mean of residual statistic Q, S represents the sample variance of Q, and $g\chi_{h,\alpha}^2$ is the threshold of χ^2 variables with scale factor g and free degree h.

3. TPLS Monitoring Model

3.1. TPLS. PLS algorithm uses two variable spaces to describe process change. However, the main component of samples contains the part which is orthogonal to $\mathbf{Y}$, and this part cannot reflect the variations related to $\mathbf{Y}$. On the other hand, PLS decomposition structure makes the residual in $\mathbf{X}$ remain very large, which is not suitable to be monitored by index Q. Therefore, Li et al. proposed a kind of total projection algorithm [12], which is based on traditional PLS decomposition. The original latent variable space is decomposed into one

subspace relevant to quality variables directly and another subspace orthogonal to quality variables. At the same time, the residual space is decomposed into subspaces with large variance and residual subspace containing noise only, using the PCA orthogonal projection technique.

By further decomposition, we can model $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\mathbf{X} = \mathbf{X}_y + \mathbf{X}_o + \mathbf{X}_r + \mathbf{E}_r,$$
$$\mathbf{Y} = \mathbf{T}_y\mathbf{Q}_y^T + \mathbf{F}, \tag{10}$$

where $\mathbf{X}_y = \mathbf{T}_y\mathbf{P}_y^T$, $\mathbf{X}_o = \mathbf{T}_o\mathbf{P}_o^T$, and $\mathbf{X}_r = \mathbf{T}_r\mathbf{P}_r^T$. $\mathbf{X}_y$ stands for the part which is relevant to $\mathbf{Y}$ directly in $\widehat{\mathbf{X}}$, $\mathbf{X}_o$ stands for the part which is orthogonal to $\mathbf{Y}$ in $\widehat{\mathbf{X}}$, and $\mathbf{X}_r$ stands for the part with large variance component in $\mathbf{E}$.

At the same time, based on the structure of PLS projection, Li et al. also performed a detailed analysis of the space structure of TPLS and drew a good conclusion [12]. Similar to PLS, TPLS also exhibits an oblique projection, but TPLS projects $\mathbf{x}$ to four different spaces, which reflect different relationship among quality variables.

For a new measurement of sample $\mathbf{x}_{\text{new}}$, the corresponding score and residual part can be calculated as follows [12]:

$$\mathbf{t}_{y\text{new}} = \mathbf{Q}_y^T\mathbf{Q}\mathbf{R}^T\mathbf{x}_{\text{new}},$$
$$\mathbf{t}_{o\text{new}} = \mathbf{P}_o^T\left(\mathbf{PR}^T - \mathbf{P}_y\mathbf{R}_y^T\right)\mathbf{x}_{\text{new}},$$
$$\mathbf{t}_{r\text{new}} = \mathbf{P}_r^T\left(\mathbf{I} - \mathbf{PR}^T\right)\mathbf{x}_{\text{new}},$$
$$\mathbf{x}_{rr\text{new}} = \left(\mathbf{I} - \mathbf{P}_r\mathbf{P}_r^T\right)\left(\mathbf{I} - \mathbf{PR}^T\right)\mathbf{x}_{\text{new}}. \tag{11}$$

Compared with PLS, TPLS model is easy to be explained and suitable for process monitoring. Similar to PLS in monitoring strategy, TPLS uses two statistic indexes T^2 and Q in process monitoring. In TPLS, $\mathbf{X}_y$, $\mathbf{X}_o$, and $\mathbf{X}_r$ represent the main variation in the process, and thus they are suitable for T^2 statistic, and $\mathbf{E}_r$ represents the residual part of the process which is suitable to be monitored by using statistic Q.

3.2. The Covariance Description of TPLS. The four spaces in TPLS can get a more detailed description of the different relationships between $\mathbf{X}$ and quality variables $\mathbf{Y}$. Based on the covariance matrix of the PLS model, the parameter matrices $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{W}$ will be obtained. Then, parameter matrix $\mathbf{R}$ is calculated by $\mathbf{R} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}$.

Combining with the covariance description form of PCA and PLS model, we can do space decomposition in the following form.

In PCA decomposition of $\widehat{\mathbf{Y}}$, characteristic vectors of the covariance matrix $\mathbf{\Sigma}_{\widehat{\mathbf{Y}}}$ are extracted to construct $\mathbf{Q}_y$. $\mathbf{\Sigma}_{\widehat{\mathbf{Y}}}$ can be expressed as

$$\mathbf{\Sigma}_{\widehat{\mathbf{Y}}} = \frac{1}{N-1}\widehat{\mathbf{Y}}^T\widehat{\mathbf{Y}} = \mathbf{Q}\mathbf{R}^T\mathbf{\Sigma_X}\mathbf{R}\mathbf{Q}^T. \tag{12}$$

Similarly, in PCA decomposition of $\widehat{\mathbf{X}}_o$ and $\mathbf{E}$, we can extract characteristic vectors of each covariance matrix to

form a loading matrix in corresponding space. Covariance matrices can be expressed as

$$
\Sigma_{\widehat{\mathbf{X}}_o} = \frac{1}{N-1} \left(\widehat{\mathbf{X}} - \mathbf{T}_y \mathbf{P}_y^T \right)^T \left(\widehat{\mathbf{X}} - \mathbf{T}_y \mathbf{P}_y^T \right)
$$

$$
= \Sigma_{\widehat{\mathbf{X}}} - \Sigma_{\widehat{\mathbf{X}}\widehat{\mathbf{Y}}} \mathbf{Q}_y \mathbf{P}_y^T - \mathbf{P}_y \mathbf{Q}_y^T \Sigma_{\widehat{\mathbf{Y}}\widehat{\mathbf{X}}} + \mathbf{P}_y \mathbf{Q}_y^T \Sigma_{\widehat{\mathbf{Y}}} \mathbf{Q}_y \mathbf{P}_y^T,
$$

$$
\Sigma_{\mathbf{E}} = \frac{1}{N-1} \left(\mathbf{X} - \widehat{\mathbf{X}} \right)^T \left(\mathbf{X} - \widehat{\mathbf{X}} \right) \tag{13}
$$

$$
= \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}} \mathbf{R} \mathbf{P}^T - \mathbf{P} \mathbf{R}^T \Sigma_{\mathbf{X}} + \Sigma_{\widehat{\mathbf{X}}},
$$

where

$$
\Sigma_{\widehat{\mathbf{X}}} = \mathbf{P} \mathbf{R}^T \Sigma_{\mathbf{X}} \mathbf{R} \mathbf{P}^T,
$$

$$
\Sigma_{\widehat{\mathbf{X}}\widehat{\mathbf{Y}}} = \mathbf{P} \mathbf{R}^T \Sigma_{\mathbf{X}} \mathbf{R} \mathbf{Q}^T, \tag{14}
$$

$$
\Sigma_{\widehat{\mathbf{Y}}\widehat{\mathbf{X}}} = \mathbf{Q} \mathbf{R}^T \Sigma_{\mathbf{X}} \mathbf{R} \mathbf{P}^T.
$$

According to the score and the residual structure model of new measurement samples, let

$$
\mathbf{R}_y^T = \mathbf{Q}_y^T \mathbf{Q} \mathbf{R}^T,
$$

$$
\mathbf{R}_o^T = \mathbf{P}_o^T \left(\mathbf{P} \mathbf{R}^T - \mathbf{P}_y \mathbf{R}_y^T \right), \tag{15}
$$

$$
\mathbf{R}_r^T = \mathbf{P}_r^T \left(\mathbf{I} - \mathbf{P} \mathbf{R}^T \right).
$$

It can be easily proved that this form is equivalent to the standard one.

The following part shows the calculation process of TPLS model using covariance information.

Covariance Description Form of TPLS Algorithm. Obtain $\Sigma_{\mathbf{X}}$ and $\Sigma_{\mathbf{XY}}$:

(1) Use GMM-PLS algorithm, and obtain parameter matrix: $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_A] \in R^{m*A}$, $\mathbf{W} = [w_1, \ldots, w_A] \in R^{m*A}$, $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_A] \in R^{l*A}$, $\mathbf{R} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}$.

(2) Calculate PCA decomposition of $\widehat{\mathbf{Y}}$: do an eigenvalue decomposition on $\Sigma_{\widehat{\mathbf{Y}}}$; obtain the loading matrix $\mathbf{Q}_y = [\mathbf{q}_{y1}, \ldots, \mathbf{q}_{yA_y}] \in R^{m*A_y}$ and principal component number: $A_y = \mathrm{rank}(\mathbf{Q})$.

(3) $\mathbf{P}_y^T = (\mathbf{Q}_y^T \Sigma_{\widehat{\mathbf{Y}}} \mathbf{Q}_y)^{-1} \mathbf{Q}_y^T \Sigma_{\widehat{\mathbf{Y}}\widehat{\mathbf{X}}}$.

(4) Calculate PCA decomposition of $\widehat{\mathbf{X}}_o$: do an eigenvalue decomposition on $\Sigma_{\widehat{\mathbf{X}}_o}$; obtain the loading matrix $\mathbf{P}_o = [\mathbf{p}_{o1}, \ldots, \mathbf{p}_{oA_o}] \in R^{m*A_o}$ and principal component number: $A_o = A - A_y$.

(5) Calculate PCA decomposition of $\mathbf{E} = \mathbf{X} - \widehat{\mathbf{X}}$: do an eigenvalue decomposition on $\Sigma_{\mathbf{E}}$; obtain loading matrix $\mathbf{P}_r = [\mathbf{p}_{r1}, \ldots, \mathbf{p}_{rA_r}] \in R^{m*A_r}$ and principal component number A_r: based on the PCA method.

4. Multimode Process Monitoring and Fault Diagnosis

4.1. Mode Division of Principal Components. According to industrial process data with the characters of multimode, we need to determine a mixed model based on historical data firstly and then design a monitoring framework. Considering covariance information required for the statistical model, multimode modeling data can be processed by GMM. It is the assumption that data are made up of different Gaussian distributions. That is, for any sample data $\mathbf{x}$, it is possible to take a certain probability from K different Gaussian distributions. As a result, global probability distribution can be expressed by the mixed model of the K Gaussian elements. It can be expressed as

$$
p(\mathbf{x} \mid \boldsymbol{\theta}) = \sum_{i=1}^{K} w_i p(\mathbf{x} \mid \boldsymbol{\theta}_i), \tag{16}
$$

where K is the number of mixture components, w_i denotes the weight of the ith Gaussian component, and $\sum_{i=1}^{K} w_i = 1$, $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \Sigma_i\}$ represents the statistical parameters. Parameters estimation usually adopts EM iterative algorithm. The corresponding multivariate Gaussian density function for the ith component is given by

$$
p(\mathbf{x} \mid \boldsymbol{\theta}_i)
$$
$$
= \frac{1}{(2\pi)^{m/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]. \tag{17}
$$

According to the rule of Bayes inference, the posterior probability of $\mathbf{x}$ belonging to the ith Gaussian component is

$$
p(\boldsymbol{\theta}_i \mid \mathbf{x}) = \frac{w_i p(\mathbf{x} \mid \boldsymbol{\theta}_i)}{\sum_{k=1}^{K} w_k p(\mathbf{x} \mid \boldsymbol{\theta}_k)}. \tag{18}
$$

However, due to factors such as production flow, batch, and specification, the quality variables of the final products have some certain degree of difference in real production processes. It may be the root cause that process data is with multimode and multistage features. Therefore, considering that the PLS algorithm is with the space decomposition under guidance of quality variables, this paper first performs mode division with principal component space $\mathbf{T}$ and acquires the mode label C_k of $\mathbf{t}_i$. This method carried out with the projection of training data can highlight the influence of quality variables better.

Based on advantages of GMM in processing multimode problems, we deal with principal components matrix $\mathbf{T}$ with GMM for acquiring $\sum_{i=1}^{K} w_{ti} = 1$ and $\boldsymbol{\theta}_{ti} = \{\boldsymbol{\mu}_{ti}, \Sigma_{ti}\}$. The total number of estimated parameters is $K((1/2)A^2 + (3/2)A + 1) - 1$, where A is the number of the principal components. Usually, A is far less than process variables number m, which can reduce the number of estimated parameters greatly and speed up the calculation.

After mode division, principal component space model based on GMM is established, where each Gauss component corresponds to different mode characteristics. For training

samples, $\mathbf{x}_i$ can be divided into the modes whose principal variable belongs to

$$C_k = P\left(\boldsymbol{\theta}_{tk} \mid \mathbf{t}_i\right) = \frac{w_{tk} p\left(\mathbf{t}_i \mid \boldsymbol{\theta}_{tk}\right)}{\sum_{i=1}^{K} w_{ti} p\left(\mathbf{t}_i \mid \boldsymbol{\theta}_{ti}\right)}. \tag{19}$$

Taking process variables $x \in R^m$ and output variables $\mathbf{y} \in \mathbf{R}^l$ into account, we construct a new vector $\mathbf{z}$ which stands for the process information as follows:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathbf{R}^{(m+l)}. \tag{20}$$

Assuming that variable $\mathbf{z}$ is satisfied with mixed Gauss distribution, the distribution parameters $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \Sigma_i\}$ can be acquired by mean($\mathbf{z}$) and cov($\mathbf{z}$) directly; prior probabilities ω_i are the same as principal space distribution ω_{ti}.

$$\Theta_{\text{GMM}} = \left\{\left\{w_1, \boldsymbol{\mu}_z^{(1)}, \Sigma_z^{(1)}\right\}, \ldots, \left\{w_K, \boldsymbol{\mu}_z^{(K)}, \Sigma_z^{(K)}\right\}\right\}. \tag{21}$$

Divide $\boldsymbol{\mu}_z^{(k)}$ and $\Sigma_z^{(k)}$ into the forms of [11]

$$\boldsymbol{\mu}_z^{(k)} = \begin{bmatrix} \boldsymbol{\mu}_x^{(k)} \\ \boldsymbol{\mu}_y^{(k)} \end{bmatrix},$$

$$\Sigma_z^{(k)} = \begin{bmatrix} \Sigma_x^{(k)} & \Sigma_{xy}^{(k)} \\ \Sigma_{yx}^{(k)} & \Sigma_y^{(k)} \end{bmatrix}. \tag{22}$$

As above, it can be noted that mode classification will be under the guidance of quality variables. Then, because the number of principal components is far less than that of process variables, this has a great advantage in the treatment of estimated parameters calculation. In addition, after the mode division of original training data, multimode information such as covariance matrices can be directly calculated, which reduce the amount of calculation and improve calculation accuracy.

4.2. Multimode TPLS Based Fault Detection. According to the principle of building PLS and TPLS, the essence is to use data information, variance, and covariance to represent process characteristics. As far as PLS is concerned, the modeling process is to maximize the covariance of linear combinations of process variables and quality variables, so the modeling process can be converted into a covariance form through the initial data $\mathbf{X}$ and $\mathbf{Y}$. Therefore, in order to adapt to the multimode characteristic of industrial process data better, we can extend multivariate statistical methods to multimode scope by covariance strategy which will improve the performance of the monitoring model.

Based on the above analysis, we can make a rational division of training data to obtain the multimode information in the process of fault monitoring. When the sample is collected and is ready for being monitored, it can be divided into corresponding models with the probability, using Bayes classification ability under the data pretreatment. Then, we can calculate the monitoring statistic of the sample to justify

which mode it belongs to. We treat the posterior probability of the monitoring sample belonging to each Gauss component as the membership degree of the corresponding model.

By using data information of probability w_i and parameters $\boldsymbol{\theta}$ to monitor the process, the comprehensive monitoring index is constructed, which can be used to monitor the fault reasonably.

For a new monitoring sample $\mathbf{x}_{\text{new}} \in \mathbf{R}^m$, the probability of sample data belonging to different modes is $P(\boldsymbol{\theta}_k \mid \mathbf{x}_{\text{new}})$.

4.3. Comprehensive Monitoring Index. According to PCA decomposition of $\widehat{\mathbf{Y}}$ in TPLS, $\mathbf{T}_y = \widehat{\mathbf{Y}} \mathbf{Q}_y$, the covariance matrix of principal components in space $\mathbf{X}_y$ can be expressed as [13]

$$\Lambda_y = \frac{1}{N-1} \mathbf{T}_y^T \mathbf{T}_y = \frac{1}{N-1} \mathbf{Q}_y^T \widehat{\mathbf{Y}}^T \widehat{\mathbf{Y}} \mathbf{Q}_y = \mathbf{Q}_y^T \Sigma_{\widehat{\mathbf{Y}}} \mathbf{Q}_y. \tag{23}$$

Available by $\widehat{\mathbf{Y}} = \mathbf{TQ}^T$ and $\mathbf{T} = \mathbf{XR}$,

$$\Lambda_y = \mathbf{Q}_y^T \mathbf{Q} \mathbf{R}^T \Sigma_{\mathbf{X}} \mathbf{R} \mathbf{Q}^T \mathbf{Q}_y = \mathbf{R}_y^T \Sigma_{\mathbf{X}} \mathbf{R}_y. \tag{24}$$

In the same way, the covariance matrices of principal components in spaces $\mathbf{X}_o$ and $\mathbf{X}_r$ can be done as in the above proof:

$$\Lambda_o = \mathbf{R}_o^T \Sigma_{\mathbf{X}} \mathbf{R}_o,$$

$$\Lambda_r = \mathbf{R}_r^T \Sigma_{\mathbf{X}} \mathbf{R}_r. \tag{25}$$

In order to realize the multimode fault monitoring, the monitoring index based on the MTPLS model is obtained by using the probability information and Bayesian inference:

$$T_{y\text{new}(k)}^2$$
$$= \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right)^T \mathbf{R}_{y(k)} \Lambda_{y(k)}^{-1} \mathbf{R}_{y(k)}^T \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right),$$
$$Q_{r\text{new}(k)} = \left\| \mathbf{x}_{rr\text{new}} \right\|^2$$
$$= \left\| \left(\mathbf{I} - \mathbf{P}_{r(k)} \mathbf{P}_{r(k)}^T\right) \left(\mathbf{I} - \mathbf{P}_{(k)} \mathbf{R}_{(k)}^T\right) \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right) \right\|^2. \tag{26}$$

Similarly,

$$T_{o\text{new}(k)}^2$$
$$= \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right)^T \mathbf{R}_{o(k)} \Lambda_{o(k)}^{-1} \mathbf{R}_{o(k)}^T \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right),$$
$$T_{r\text{new}(k)}^2$$
$$= \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right)^T \mathbf{R}_{r(k)} \Lambda_{r(k)}^{-1} \mathbf{R}_{r(k)}^T \left(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_x^{(k)}\right). \tag{27}$$

The threshold can be inferred by the setting in standard TPLS.

In summary, we make use of covariance information mainly to calculate and then to achieve process monitoring in MTPLS. Compared with standard TPLS, the covariance model is more suitable for monitoring multimode processes and making full use of data information in the process of model construction and fault monitoring. Avoiding direct classification on data, the covariance model reduces the effect of classification on the final performance monitoring of the process.

4.4. Quality-Related Combined Index. In TPLS based process monitoring, space $\mathbf{X}_y$ represents the change part related to quality variable, while space $\mathbf{E}_r$ represents the uncertain parts related to quality variable. They reflect two different kinds of quality-related faults. Therefore, it is necessary to observe two subspaces at the same time. In practice, a unified monitoring index is more popular than the two separate ones. In PCA based fault detection, Yue and Qin proposed a combined index [14]. Li et al. proposed a combined one for TPLS based process monitoring [12]. Similarly, a combined index which incorporates T_y^2 and Q_r is proposed in a way as follows:

$$\varphi_y = \frac{T_y^2}{\delta_y} + \frac{Q_r}{\delta_{rr}} = \mathbf{x}^T \mathbf{\Phi} \mathbf{x}, \tag{28}$$

where

$$\mathbf{\Phi} = \frac{\mathbf{R}_y \Lambda_y^{-1} \mathbf{R}_y^T}{\delta_y} + \frac{\left(\mathbf{I} - \mathbf{RP}^T\right)\left(\mathbf{I} - \mathbf{P}_r \mathbf{P}_r^T\right)\left(\mathbf{I} - \mathbf{PR}^T\right)}{\delta_{rr}}. \tag{29}$$

ζ^2 is the threshold of this combined index which can be obtained by approximate distribution $\zeta^2 = g\chi_{h,\alpha}^2$. It is supposed that there is no fault in the process when the monitoring result is $\varphi_y < \zeta^2$.

Scale factor g and free degree h are calculated in

$$g = \frac{\text{tr}\,(\mathbf{S}\mathbf{\Phi})^2}{\text{tr}\,(\mathbf{S}\mathbf{\Phi})},$$

$$h = \frac{[\text{tr}\,(\mathbf{S}\mathbf{\Phi})]^2}{\text{tr}\,(\mathbf{S}\mathbf{\Phi})^2}, \tag{30}$$

where $\mathbf{S} = \text{cov}(\mathbf{x}) = \Sigma_{\mathbf{X}}$, which is the covariance matrix of process variable $\mathbf{x}$. Using this combined index, we can simultaneously monitor the anomalies in the two subspaces and thus monitor the faults associated with the quality variables $\mathbf{Y}$.

4.5. Gradient Contribution Rate for Fault Diagnosis. It is necessary to isolate the faulty variables after a fault is detected. As a common fault separation method, the contribution plot assumes that the variables which have greater contribution to the monitoring statistics are very likely to be faulty variables. According to the description framework of complete decomposition of contribution proposed by Alcala and Qin, contribution to the combined index can be described as the following form [15]:

$$\text{con}\,\mathbf{t}_i = \left(\gamma_i \mathbf{x}\right)^2 = \left(\xi_i^T \mathbf{\Phi}^{1/2} \mathbf{x}\right)^2,$$

$$\text{index} = \left\|\mathbf{\Phi}^{1/2} \mathbf{x}\right\|^2 = \sum_{i=1}^m \text{con}\,\mathbf{t}_i, \tag{31}$$

where γ_i represents the ith row of matrix $\mathbf{\Phi}^{1/2}$, ξ_i represents the ith row of identity matrix, and m represents the number of variables in one sample.

Traditional contribution plot method is used for analyzing a specific sample when the fault is detected, which shows the contribution value of each variable to one monitoring index in bar chart. After that, the variables with greater contribution will be selected as the possible cause of fault. Westerhuis et al. put forward a generalized contribution to statistics form and a method of obtaining the control limits for variable contributions [16]. Choi et al. proposed specific statistical methods to set the upper limit of the variable contribution to the four monitoring statistics [7]. Li et al. proposed a kind of contribution plot based on TPLS, which describes the contribution of all variables to monitoring index T_y^2 and Q_r in a unified way [12].

For the fault diagnosis method based on traditional contribution figure for one single sample after fault occurrence, there are some flaws that cannot well describe fault source and the change of other malfunction variables caused by fault source. In order to combine the idea of analyzing the contribution rate of faulty variables along the time coordinates with the change of the variable itself, reducing the impact of variable magnitude of value on the contribution rate, we refer to the gradient contribution rate to solve the fault variable analysis.

First, we introduce a mathematical symbol $\odot$ and a scale factor vector $\mathbf{v} = [v_1, \ldots, v_m]^T$, where $\odot$ indicates element product. $\mathbf{x} \odot \mathbf{v} = [\mathbf{x}_1 v_1, \ldots, \mathbf{x}_i v_i, \ldots, \mathbf{x}_m v_m]^T$, and $\mathbf{x}_i v_i$ indicates the change of variable $\mathbf{x}_i$. As can be segmented, if $v_i > 1$, then $|\mathbf{x}_i v_i| > |\mathbf{x}_i|$; if $v_i = 1$, then $\mathbf{x}_i v_i = \mathbf{x}_i$; if $v_i < 1$, then $|\mathbf{x}_i v_i| < |\mathbf{x}_i|$. So, equation $\varphi(\mathbf{v} \odot \mathbf{x})|_{\mathbf{v}=\mathbf{1}_m} = \varphi(\mathbf{x})$ can be established.

It can be seen from the first-order Taylor series expansion of $\varphi(\mathbf{v} \odot \mathbf{x})$ near $\mathbf{v} = \mathbf{1}_m$ that

$$\varphi(\mathbf{x} \odot \mathbf{v}) \approx \varphi(\mathbf{x}) + \sum_{i=1}^m \left.\frac{\partial \varphi(\mathbf{x} \odot \mathbf{v})}{\partial v_i}\right|_{\mathbf{v}=\mathbf{1}_m} (v_i - 1). \tag{32}$$

Based on the above conclusion, the contribution rate may be defined as follows.

For a monitoring sample $\mathbf{x}$, $C(\mathbf{x}, i) \triangleq |(\partial \varphi(\mathbf{x} \odot \mathbf{v})/\partial v_i)|_{\mathbf{v}=\mathbf{1}_m}|$ indicates the contribution rate of the ith variable to index φ.

As described above, the contribution rate represents the gradient of each variable to detection index under the same abnormal changes. Variables which are with great contribution will be considered with great influence to index φ, the same to quality variable.

For a new monitoring sample $\mathbf{x}_{\text{new}}$, the contribution rate of the ith variable can be calculated as

$$\left|\left.\frac{\partial \varphi(\mathbf{x}_{\text{new}} \odot \mathbf{v})}{\partial v_i}\right|_{\mathbf{v}=\mathbf{1}_m}\right|. \tag{33}$$

As a result, the gradient contribution rate based on comprehensive monitoring index φ_y can be expressed as follows:

$$C(\mathbf{x}_{\text{new}}, i) = \left|\left.\frac{\partial \left(\xi_i^T \mathbf{\Phi}^{1/2} (\mathbf{x}_{\text{new}} \odot \mathbf{v})\right)^2}{\partial v_i}\right|_{\mathbf{v}=\mathbf{1}_m}\right|$$

$$= \left|2\xi_i^T \mathbf{\Phi}^{1/2} \mathbf{x}_{\text{new}} \mathbf{x}_{\text{new},i}\right|, \tag{34}$$

where $\mathbf{x}_{\text{new},i}$ represents the value of the ith variable in monitoring sample $\mathbf{x}_{\text{new}}$.
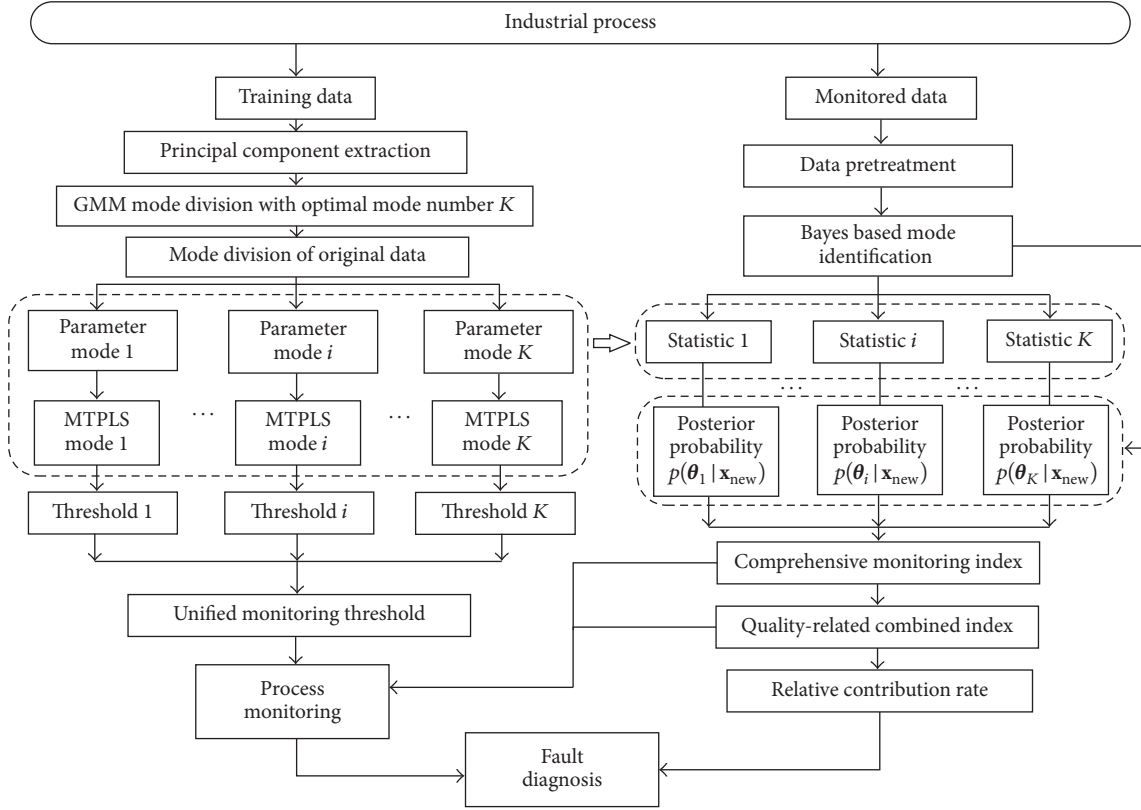
FIGURE 1: Schematic diagram of the proposed PLS-MTPLS and Bayesian-based process monitoring and diagnosis method.

Due to the diffusion effect of fault, the method of setting absolute control limits using absolute value of variable contribution for fault diagnosis is not with good effect. Therefore, we use relative contribution rate; namely,

$$C_r\left(\mathbf{x}_{\text{new}}, i\right) = \frac{C\left(\mathbf{x}_{\text{new}}, i\right)}{\sum_{i=1}^{m} C\left(\mathbf{x}_{\text{new}}, i\right)}, \tag{35}$$

where relative contribution rate satisfies

$$\sum_{i=1}^{m} C_r\left(\mathbf{x}_{\text{new}}, i\right) = 1. \tag{36}$$

As described above, in index φ_y based quality-related fault diagnosis, the contribution rate can reflect contribution gradients of variables to the monitoring index. Therefore, those variables which have a larger contribution rate are able to affect combined index and quality variables significantly.

4.6. Framework of Fault Detection and Diagnosis. The schematic diagram of the proposed process monitoring and diagnosis is shown in Figure 1. Detailed procedures for multimode process detection can be summarized below:

(1) Collect a set of historical training data under all possible operating modes and determine the number of modes.

(2) Use EM algorithm to learn the Gaussian mixture model of principal component space and estimate the model parameter set $\mathbf{\Theta_T}$ based on the iterative steps.

(3) Do multimode division and multimode information acquisition of process data according to C_k. Then, for each monitored sample $\mathbf{x}_{\text{new}}$, compute its posterior probabilities belonging to all Gaussian components through Bayesian inference strategy.

(4) Calculate local monitoring statistics for the monitored sample $\mathbf{x}_{\text{new}}$ within each Gaussian component and integrate them into the comprehensive index with probabilities.

(5) Integrate the quality-related monitoring statistics into a quality-related combined index φ_y.

(6) Specify a confidence level $(1-\alpha)100\%$ for determining control threshold ζ^2 and generate the monitoring plot for all the monitored samples.

(7) Detect the abnormal operating condition at the monitored samples satisfying $\varphi_y > \zeta^2$ which is helpful for fault diagnosis.

(8) Calculate the relative contribution rate of variables to the combined index φ_y before and after fault occurrence and generate the contribution rate plot for fault diagnosis analysis.

5. Application to HSMP

5.1. Hot Strip Mill Process. HSMP (hot strip mill process) is an extremely complex industrial production process. In
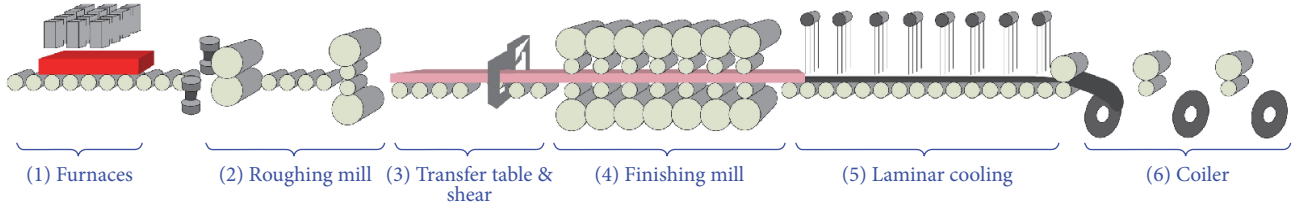
(1) Furnaces   (2) Roughing mill   (3) Transfer table & shear   (4) Finishing mill   (5) Laminar cooling   (6) Coiler
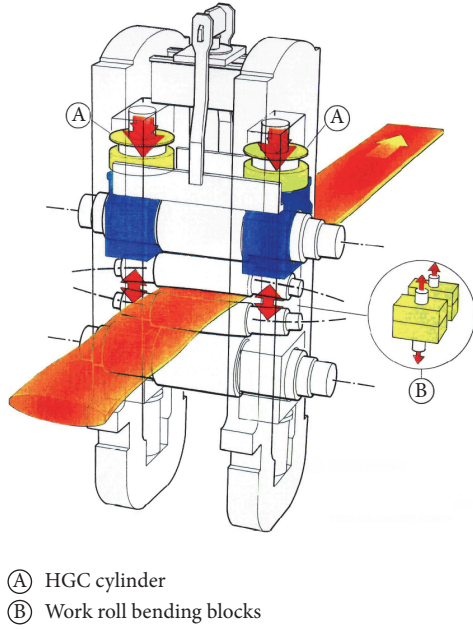

(A) HGC cylinder
(B) Work roll bending blocks

FIGURE 3: The structure of mill stand.

the process of production, improving the quality of products can bring about higher economic and social benefits for the factory. Typical HSM machine production line is mainly composed of reheating furnace, roughing mill, transfer table, crop shear, finishing mill group, run-out table cooling, and coiler. Figure 2 shows the whole process flow chart. The reheating furnace can ensure the temperature of the strip reaches 1200 degrees Celsius before roughing mill. A slab of thickness of 100~200 mm is sent to the roughing mill group after cutting off the scales, eventually forming 28~45 mm thick middle slab through several times of rolling. Through the transport of transfer table and in turn with insulation cover, crop shear, and high pressure water descaling, the slab runs into seven stands of finishing mill group. In order to enhance the performance of the final product, the steel plate needs to go through laminar cooling. This paper focuses on fault monitoring in the part of finishing mill process (FMP).

As shown in Figure 2, FMP consists of seven stands. Every stand contains two working rolls and backup rolls, which are driven by their own power drive units. The distance between two working rolls is called roll gap, which can be adjusted by the hydraulic device. A detailed structure diagram of the finishing roll is shown in Figure 3. This means that the strip

will go through all the seven stands during the finishing mill process.

In whole FMP, it is noted that the stands are actually not working independently but are coupled with each other by different control schemes. The thickness in the exit of the last stand is the key factor which directly affects the quality of products. Whole finishing mill process is controlled by automatic thickness control system. It can be seen that there is an obvious hysteresis control of the exit thickness. Not until the abnormal value of the exit thickness is detected, caused by some fault of front stands, can the thickness control system be started. Therefore, establishing real-time acquisition of the relationship between the process variables and exit thickness and then monitoring the thickness by real-time measuring process variables become very meaningful [14].

5.2. Fault Detection Simulation Analysis. Production specification can be determined by different thicknesses of the steel strip in HSMP which should meet different industrial demands. We select the steel plate data of two specifications for modeling: one is the thickness of 2.70 mm and the other is 3.95 mm. The sampling interval for the variable is 0.01 s and 4000 samples are used for training modeling.

In the actual finishing mill process, we can collect the data information including roll gap, milling force between working rolls, and bending force in every stand. Generally speaking, the exit thickness has more relationships with roll gap and milling force than with bending force. Using data collected under normal operating conditions, GMM iterative learning is performed in principal component space which is under the guidance of quality variables. With the model division result, the process of multimode parameters calculation of the original data is carried out. Figure 4 shows the clustering distribution of two kinds of normal production. In this part, the clustering numbers are $K1 = 3$ and $K2 = 5$ which are fixed according to Yu et al. The K-means algorithm is applied to roughly calculate $\boldsymbol{\mu}_i^k$. Randomly initialize the value of $\boldsymbol{\Sigma}_i^k$ before GMM iterative learning. Then, we establish the proposed PLS-MTPLS model. Variables concerned in FMP are as follows.

Process Variables

$\mathbf{x}_1 \sim \mathbf{x}_7$: average gap of F_i stand, $i = 1, \ldots, 7$, μm

$\mathbf{x}_8 \sim \mathbf{x}_{14}$: the force between supporting and working roll in F_i stand, $i = 1, \ldots, 7$, KN

(a) $K1 = 3$

(b) $K2 = 5$
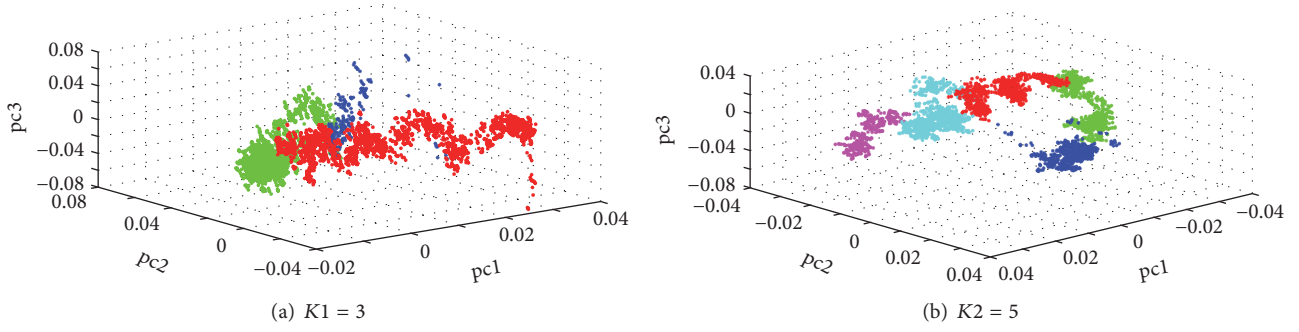
FIGURE 4: Principal components clustering distribution where $K1 = 3$ and $K2 = 5$ (pc: principal component).
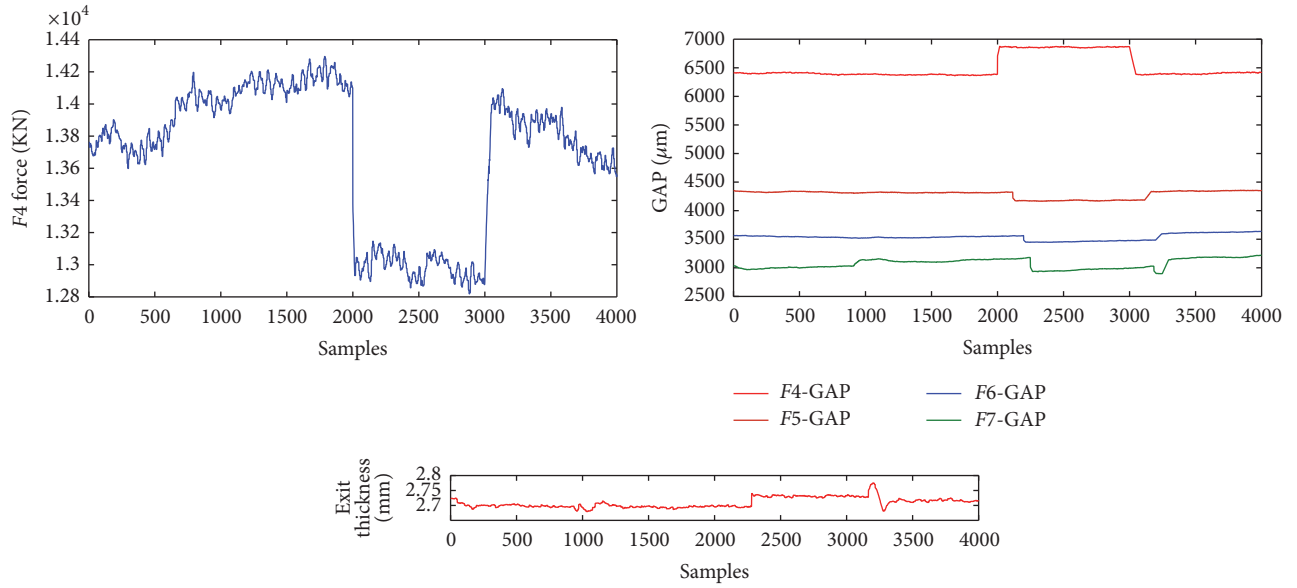
FIGURE 5: Curve of fault source $\mathbf{x}_4$ and affected variables.

$\mathbf{x}_{15} \sim \mathbf{x}_{20}$: the bending force in the working roll of F_i stand, $i = 1, \ldots, 7$, KN

Quality Variable

$\mathbf{y}$: thickness of strip at the exit of FMP, mm

For different types of faults that may occur in FMP, we select three encountered faults as a detected object in this section which are shown in Table 1.

According to the exit thickness value of the strip steel under three types of fault condition, it is obvious that fault 1 and fault 3 are quality-related, while fault 2 is quality-unrelated. As Figures 6 and 9 show, the method based on PLS-MTPLS gives a higher fault detection rate for fault 1 and fault 3. And for the quality-unrelated fault 2, PLS-MTPLS inherits the effect of space division in traditional TPLS method, making the monitoring index T_y^2 which is directly related to the quality have a relatively low rate of false alarm.

To examine the advantages of our proposed approach, a comparison research has been done using two evaluating indexes: FDR and FAR.

FAR

$$= \frac{\text{Number of samples} \left(\varphi_y > \zeta^2 \mid \text{quality is normal} \right)}{\text{total fault} - \text{free samples}},$$

FDR

$$= \frac{\text{Number of samples} \left(\varphi_y > \zeta^2 \mid \text{quality is faulty} \right)}{\text{total faulty samples}}. \tag{37}$$

False detection rates and false alarm rates are counted for three types of fault and statistical results are shown in Table 2. It shows that PLS-MTPLS method performs better.

Fault 1 represents the failure of hydraulic roll gap control structure. Fault occurs at about 20 s, namely, the 2000th monitoring sample. The values of roll gap $\mathbf{x}_4$ in the fourth
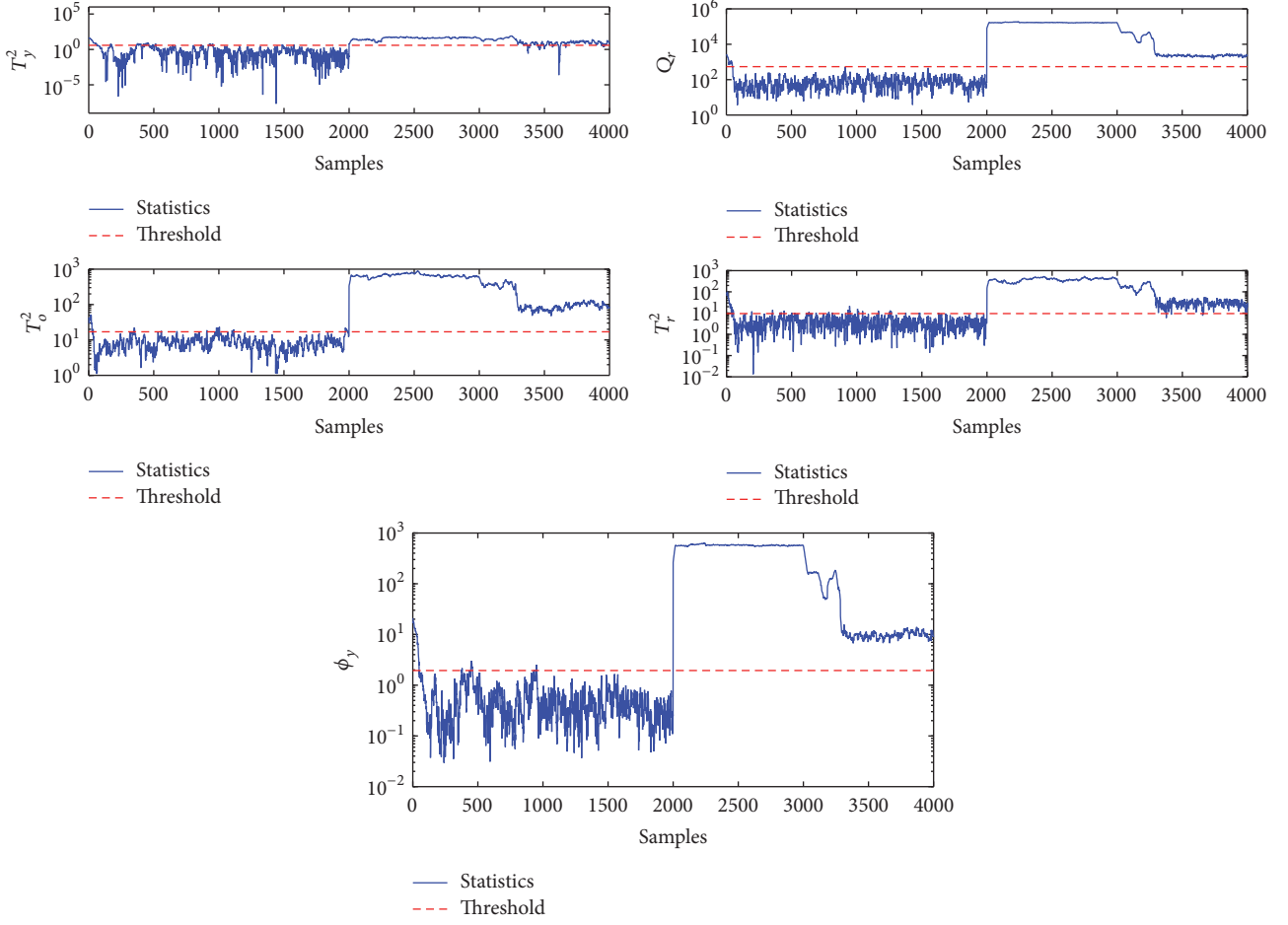
Figure 6: Detection result of fault 1.

Table 1: Faults that occurred in FMP.

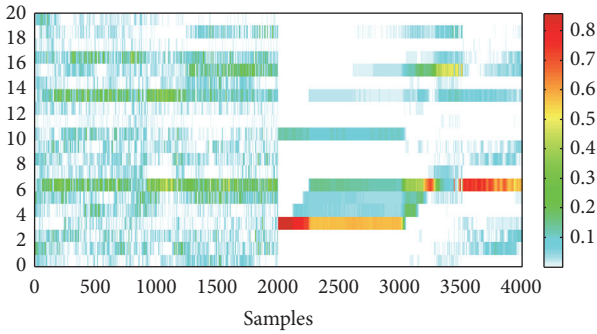| Fault number	Fault description	Duration	Type
1	Malfunction of gap control loop in $F4$ stand	20 s–30 s	Quality-related
2	Fault of roll bending force measuring sensor in $F5$ stand	10 s–20 s	Quality-unrelated
3	10% stiction of the cooling valve between $F2$ and $F3$ stands	10 s–20 s	Quality-related

Figure 7: Diagnosis result of fault 1.

stand are directly affected, and then the sampling values of milling force $\mathbf{x}_{11}$ in the fourth stand have also been affected.

Because of the influence of feedback control system, roll gaps and milling forces will be changed in the following stands, and then finally the exit thickness is affected. As shown in Figure 5, there is a delay for change of exit thickness value, namely, quality variables with respect to fault occurrence. But for fault detection results, as shown in Figure 6, there is almost no delay. From the point of view of this analysis, detection results can be a good reference for field staffs, in order to take response measurements timely.

Figure 7 gives the observation of change of relative contribution rate for fault 1. As is shown, we can clearly see that many contribution rate values of related variables have changed since the 2000th monitoring sample. When the fault is detected, according to the observation of relative contribution rate, variable $\mathbf{x}_4$ which has the largest relative contribution rate is diagnosed firstly. As a result, we can conclude that roll gap of $F4$ stand is the source of fault. At the same

TABLE 2: Detection performance comparison.

Fault number	Type	MPLS (T^2)	TPLS (φ_y)	MTPLS ($T_y^{\ 2}$)	MTPLS (φ_y)	PLS-MTPLS ($T_y^{\ 2}$)	PLS-MTPLS (φ_y)
1	FDR	0.7968	0.8977	0.8620	0.9995	0.9435	0.9995
2	FAR	0.3357	0.4335	0.1335	0.2610	0.0080	0.2605
3	FDR	0.7780	0.8510	0.8848	0.9729	0.9092	0.9735
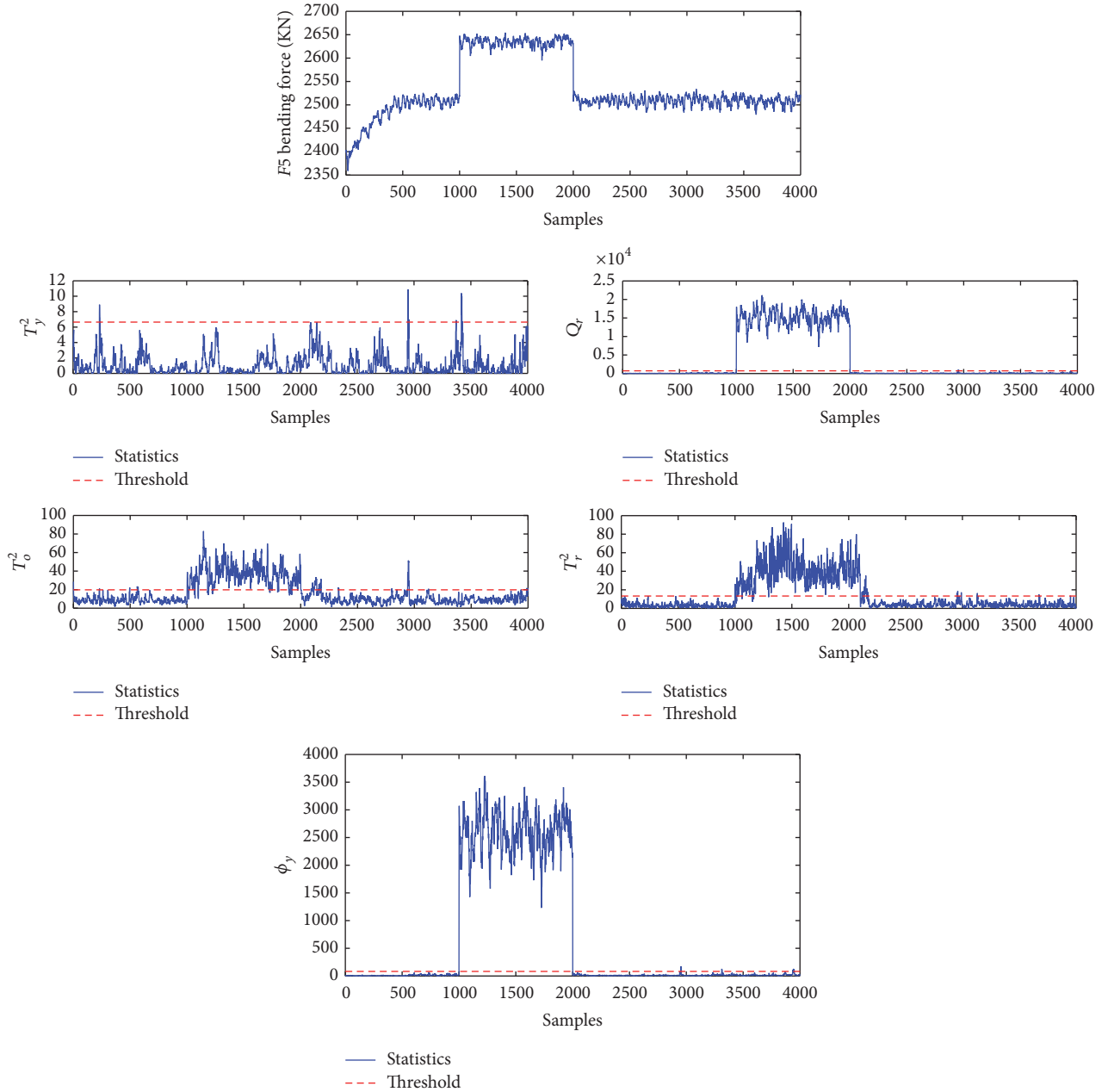
FIGURE 8: Detection result of fault 2.

time, variables $\mathbf{x}_{11}$ and $\mathbf{x}_5 \sim \mathbf{x}_7$ are subsequently diagnosed which are affected by the fault source, thereby causing fault propagation. Figure 7 shows curves of variables change in real data of fault source variable and variables affected. The diagnostic analysis is in accordance with the actual production situation. As a result, the relative contribution rate can not only

diagnose fault variables, but also show the order of fault variables transmission. Then, it can help in finding out the real source of fault with causal relationship among these variables.

Fault 2 represents the fault of sampling value of bending force in F5 stand, which is a kind of step transition. When the fault occurs, the value of variable $\mathbf{x}_{18}$ will increase greatly.
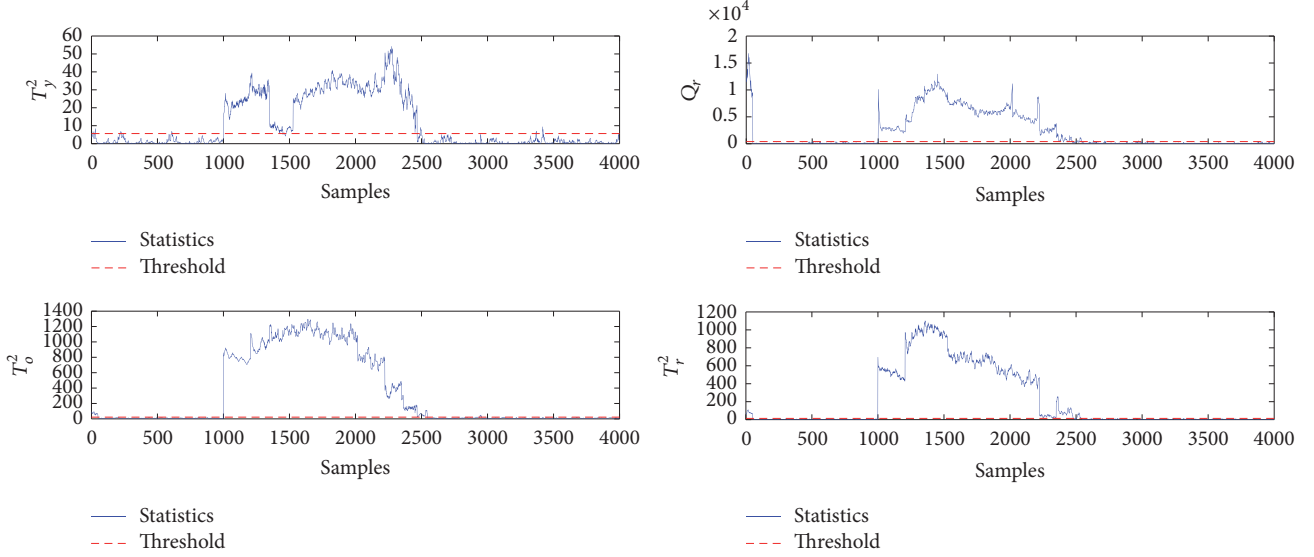
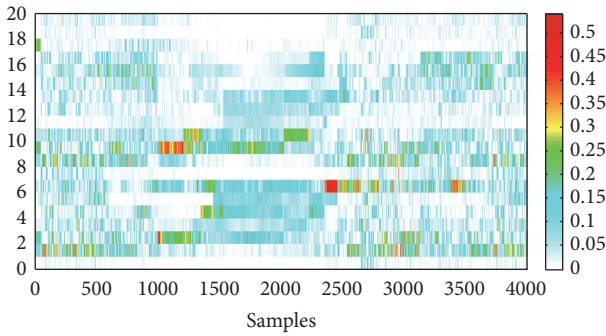FIGURE 9: Detection result of fault 3.

FIGURE 10: Diagnosis result of fault 3.

Then, with feedback regulation of automatic control system, the bending force value in $F6$ and $F7$ stands will be changed correspondingly. But this kind of fault will only cause the change of strip plate, not thickness as shown in Figure 8.

Fault 3 is a kind of fault in cooling valve between $F2$ and $F3$ stands, which is usual in the process of finishing mill. It will make the rolling force and roll gap in stands following $F3$ stand be changed. Based on the monitoring results of comprehensive index, fault 3 can be detected timely as shown in Figure 9. The change of relative contribution rate is shown in Figure 10. It can be noted that variable $\mathbf{x}_3$ is affected firstly, followed by variable $\mathbf{x}_{10}$ and others. From the above analysis, we can draw a conclusion that the fault diagnosis method based on the relative contribution rate can be applied to FMP effectively.

In this section, we focus on the research of exit thickness of the strip. Twenty variables among measured variables in FMP were selected for building PLS-MTPLS model. Based on it, a kind of comprehensive monitoring index and a kind of relative contribution rate were established for fault monitoring and diagnosis, respectively, for three common faults. Results of monitoring and diagnosis verified that

PLS-MTPLS has higher FDR and lower FAR than traditional multivariate statistics methods shown in Table 2. In addition, compared with MTPLS which clusters with original data directly, this method is with better monitoring effects in statistics T_y^2 of principal component which can be seen in Table 2.

6. Conclusion

In this paper, a new PLS-MTPLS method is proposed on the basis of covariance descriptions of PCA and PLS algorithm for multimode process monitoring. After mode division of quality-related principal components, multimode information is embedded into the monitoring model by integrating GMM with TPLS, which avoids the direct use of process training data for modeling. Based on the quality-related multimode monitoring model PLS-MTPLS, a kind of comprehensive monitoring index is applied to execute real-time online monitoring. Then, a combined index is constructed for improving monitoring efficiency and extended to fault diagnosis by relative gradient contribution rate calculation.

The efficiency and superiority of PLS-MTPLS are demonstrated through application to the monitoring of HSMP. As can be seen from the comparison and analysis, the proposed approach can reduce computational complexity and be more suitable for multimode processes.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

References

[1] L. H. Chiang, R. D. Braatz, and E. L. Russell, *Fault Detection and Diagnosis in Industrial Systems*, Springer, London, UK, 2001.

[2] D.-H. Hwang and C. Han, "Real-time monitoring for a process with multiple operating modes," *Control Engineering Practice*, vol. 7, no. 7, pp. 891–902, 1999.

[3] S. Lane, E. B. Martin, R. Kooijmans, and A. J. Morris, "Performance monitoring of a multi-product semi-batch process," *Journal of Process Control*, vol. 11, no. 1, pp. 1–11, 2001.

[4] J. Chen and J. Liu, "Mixture principal component analysis models for process monitoring," *Industrial and Engineering Chemistry Research*, vol. 38, no. 4, pp. 1478–1488, 1999.

[5] S. J. Zhao, J. Zhang, and Y. M. Xu, "Monitoring of processes with multiple operating modes through multiple principle component analysis models," *Industrial and Engineering Chemistry Research*, vol. 43, no. 22, pp. 7025–7035, 2004.

[6] X.-T. Doan and R. Srinivasan, "Online monitoring of multiphase batch processes using phase-based multivariate statistical process control," *Computers and Chemical Engineering*, vol. 32, no. 1-2, pp. 230–243, 2008.

[7] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Computers and Chemical Engineering*, vol. 28, no. 8, pp. 1377–1387, 2004.

[8] C. K. Yoo, K. Villez, I.-B. Lee, C. Rosén, and P. A. Vanrolleghem, "Multi-model statistical process monitoring and diagnosis of a sequencing batch reactor," *Biotechnology and Bioengineering*, vol. 96, no. 4, pp. 687–701, 2007.

[9] J. Yu and S. J. Qin, "Multiway gaussian mixture model based multiphase batch process monitoring," *Industrial and Engineering Chemistry Research*, vol. 48, no. 18, pp. 8585–8594, 2009.

[10] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, New York, NY, USA, 1998.

[11] K. Peng, K. Zhang, B. You, and J. Dong, "Quality-related prediction and monitoring of multi-mode processes using multiple PLS with application to an industrial hot strip mill," *Neurocomputing*, vol. 168, pp. 1094–1103, 2015.

[12] G. Li, S.-Z. Qin, Y.-D. Ji, and D.-H. Zhou, "Total PLS based contribution plots for fault diagnosis," *Acta Automatica Sinica*, vol. 35, no. 6, pp. 759–765, 2009.

[13] K. Zhang, H. Hao, Z. Chen, S. X. Ding, and K. X. Peng, "A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches," *Journal of Process Control*, vol. 33, pp. 112–126, 2015.

[14] H. H. Yue and S. J. Qin, "Reconstruction-based fault identification using a combined index," *Industrial and Engineering Chemistry Research*, vol. 40, no. 20, pp. 4403–4414, 2001.

[15] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593–1600, 2009.

[16] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, "Generalized contribution plots in multivariate statistical process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 95–114, 2000.