IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 10, 2021, accepted January 25, 2021, date of publication January 28, 2021, date of current version February 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3055226

A Novel Multi-Mode Bayesian Method for the Process Monitoring and Fault Diagnosis of Coal Mills

WEI FAN[1], SHAOJUN REN[1], QINQIN ZHU[2], (Member, IEEE), ZHIJUN JIA[3], DELONG BAI[3], AND FENGQI SI[1]

[1]Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, Southeast University, Nanjing 210096, China
[2]Department of Chemical Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[3]Inner Mongolia Jinglong Electric Power Generation Company Ltd., Fengzhen 012100, China

Corresponding author: Shaojun Ren (rsj@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51976031, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions and Qinglan Project of Jiangsu Province, China.

ABSTRACT Process monitoring and fault diagnosis (PM-FD) of coal mills are essential to the security and reliability of the coal-fired power plant. However, traditional methods have difficulties in addressing the strong nonlinearity and multi-modality of coal mills. In this paper, a novel multi-mode Bayesian PM-FD method is proposed. Gaussian mixture model (GMM) is first applied to identify the operating modes of the coal mill. Subsequently, combined with multi-output relevance vector regression (MRVR), Bayesian inference is introduced to reconstruct and monitor the newly observed samples from different running modes. Additionally, the squared prediction error and the contribution plot method are employed for fault detection and isolation. The performance of the proposed PM-FD method is verified through its application in a self-defined nonlinear system and two actual fault cases of a medium-speed coal mill. Compared with the traditional methods, the experimental results demonstrate the effectiveness of the proposed method.

INDEX TERMS Process monitoring, multi-mode fault diagnosis, coal mills, Gaussian mixture model, Bayesian multi-output relevance vector reconstruction.

I. INTRODUCTION

Coal mills are crucial equipment of the coal-fired power plant's pulverizing system, in which the raw coal is crushed and ground into coal powder. Due to the harsh conditions, coal mills may suffer from blockage, fire, shutdown, and wear of its components [1]. Once abnormal conditions or malfunctions occur, the coal flow rate as well as the quality of coal powder cannot be guaranteed, which negatively affects the combustion state. In this case, the generated superheated steam cannot meet the demand of the steam turbine. For the worst case, it may lead to furnace flame extinction and emergency shutdown of the whole unit. Therefore, effective PM-FD methods of coal mills are urgently needed for the operational security and reliability of the power plant.

Quantities of modeling approaches have been proposed for PM-FD of coal mills, which are divided separately into model-based methods, signal-based methods, and historical data-driven methods [2]. Based on basic principles and certain assumptions, model-based methods establish mathematical models to characterize coal mills. Researchers realize PM-FD of coal mills by comparing actual measurements and models' expectations [3]–[7]. However, due to the strong complexity and non-linearity of coal mills [8], establishing accurate mathematical models is by no means a simple task. In order to avoid the difficulty of modeling, some researchers proposed signal-based methods from the perspective of analyzing high frequency signals of mechanical equipment [9]. The health status of coal mills is evaluated based on the comparison of time-frequency characteristics between normal and fault operating conditions [10], [11]. When utilizing these methods, a considerable number of sensors for measuring high-frequency signals need to be installed, and the costs of installation and maintenance are relatively high [5]. Compared with the above two methods, historical process data-driven methods do not need to take basic mechanism principles into account [12]–[14].

The associate editor coordinating the review of this manuscript and approving it for publication was Fatih Emre Boran.

Instead, their main purpose is to obtain the differences between normal operation data and fault operation data. With the differences analyzed, the health status can be determined [1], [8], [15]–[17]. Hence, data-driven approaches have become increasingly popular because of their flexibility and maneuverability [18], [19], and they are widely applied in PM-FD of complex industrial processes [20].

Among the dominant data-driven methods for PM-FD, principal component analysis (PCA), artificial neural networks (ANN) and support vector machine (SVM) are three of the most prevalent algorithms. In PCA methods, for example, Qin [21], Alcala and Qin [22] and [23] proposed a reconstruction-based contributions method and analyzed its efficiency in guaranteeing correct fault diagnosis. Ge and Song [24] discussed and confirmed the effectiveness of applying distributed PCA to plant-wide process monitoring. These methods do help improve the PM-FD performance, but lack of robustness and with inadequate monitoring accuracy for complex nonlinear systems [25]. Zhang *et al.* [26] used a KPLS method for PM-FD problems and verified its effectiveness. Considering the characteristics of nonlinearity and complexity of practical industrial processes, ANN-based methods perform apparently better. Eslamloueyan [27] developed a hierarchical artificial neural network (HANN) to isolate faults of the Tennessee Eastman Process and testified its performance. Ren *et al.* [28] combined the idea of reconstruction and auto-associative neural network (AANN), and proved its efficiency in fault diagnosis. However, difficulties such as the determination of the optimum structure, local optimum, and overfitting are still remaining to be solved [29]. SVM is another effective fault diagnosis method based on statistical theory [30]. It performs well in reaching the global optimum by transforming complex issues into convex problems. Kulkarni *et al.* [31] used an invariant SVM to improve the faults classification performance. Mahadevan and Shah [32] employed SVM for fault diagnosis and proved its superiority. Although the SVM model has a simple structure and great robustness, the balance between its soft-margin error and trade-off constant still needs to be investigated. In summary, the previous works suffer from some shortcomings, among them are difficulties in handling complex nonlinear systems with PCA, local optimum and overfitting with ANN, the disadvantages of choosing parameters with SVM, and lack of multi-modal considerations.

This proposal aims to overcome the shortcomings of the earlier works and develop a novel method for dealing with the strong nonlinearity and multi-modality of coal mills. Different from existing methods, the relevance vector machine (RVM) [33] has a Bayesian framework and offers good generalization performance through sparse predictors [34]. Most parameters are set to zero during the learning process, which is effective for achieving good prediction results and avoiding over-fitting problems [33]. In addition, RVM avoids the disadvantages of determining the soft-margin error and the trade-off constant. Furtherly, RVM has been extended to multi-output relevance vector regression (MRVR) [35], [36], which has the potential of acting as a multivariate reconstruction technique for PM-FD. However, to achieve great map relationships between inputs and reconstructions with MRVR, all possible deviations are needed for training [37]. Operation data collected from coal mills are mainly under normal conditions, it is difficult to acquire all abnormal samples in the actual process. Adding certain artificial deviations to training data seems to be an alternative way but it is hard to implement, because the dimension of the variables is relatively high. Considering the limitations of applying MRVR in the PM-FD of coal mills, a reconstruction method based on Bayesian inference [38], [39] is designed in this paper, namely Bayesian multi-output relevance vector reconstruction (BMRVR). The highlight of BMRVR is that a probabilistic mapping between inputs and targets is established, and the requirement for faulty training data is eliminated. Another important aspect is that coal mills are always running at various different operating modes due to unit load fluctuations, resulting in the invalid assumption of a multivariate Gaussian distribution [43]. A change in the process running mode could make it hard to learn the underlying data structure with only one single model. Besides, the data structure of such multivariate process is quite complicated in the presence of process characteristics such as process degradation. The application of local PM-FD methods can help address these challenges by dividing the data into separate groups and modeling them separately [44]. As a result, methods for dividing the data into various regions are of great significance being preceded in applying local PM-FD methods. Gaussian mixture model (GMM) [40] provides the superior capability of dividing data samples into different modes by local Gaussian distribution, and each Gaussian component corresponds to one possible operating mode. Quantities of GMM based techniques have been applied to deal with the multimode processes and the benefits have been remarked as well [41]–[49]. Thereby GMM can serve multi-mode PM-FD of coal mills and help improve the reliability and accuracy.

Therefore, the contribution of this paper is that we develop a novel multi-mode Bayesian inference-based reconstruction method for PM-FD of coal mills. The new method can deal with the deficiency of the traditional PM-FD methods in lack of faulty training data and the invalid Gaussian distribution assumption. First of all, GMM is established to model the normal training samples, thus obtaining the prior probabilities of different operating modes of coal mills. The local MRVR algorithm is then taken for different Gaussian batches. Based on the above framework, once a new input occurs, it is first determined the operation state to which it belongs. Afterward, Bayesian inference is employed to explore possible deviations between the fault vector and normal data. The rest of this paper is organized as follows: Section II introduces the basic concepts and common faults of medium speed coal mills. In Section III, the basic knowledge of GMM, MRVR and proposed BMRVR is described. Section IV presents the detailed process of the proposed multi-mode PM-FD framework. Experiments to a self-defined nonlinear system and

a medium speed coal mill in a coal-fired power plant are discussed in Section V. Finally, conclusions are drawn in Section VI.

II. DESCRIPTION OF MEDIUM SPEED COAL MILLS

In most modern coal-fired power plants, medium speed coal mills are widely equipped because of their flexibility and economic efficiency.

A. OPERATION OF COAL MILLS

The work in this part is based on a ZGM-113N medium speed coal mill equipped in a subcritical 600MW coal-fired unit. As shown in Fig. 1, the raw coal falls into the coal feeder, enters the coal mill through the inlet pipe, and finally falls onto a constant speed rotating grinding table. After being ground, qualified coal fines are carried by the mixed primary air and sent to the furnace for combustion. Fig. 2 shows the schematic diagram of the pulverizing process inside the coal mill. The raw coal is ground by the relative motion of the roller and the bowl. On the edge of the mill, pulverized coals are dried and carried by the mixed primary air. Large and heavy particles will fall onto the grinding table, namely the first separation. Coal-air mixture particles then go through the second separation when passing the rotary separator. During this process, larger particles fall back to the grinding area along the wall because of the inertial interception and the separation function. Therefore, the coal powder sent to the furnace can meet the requirement of the coal fineness.

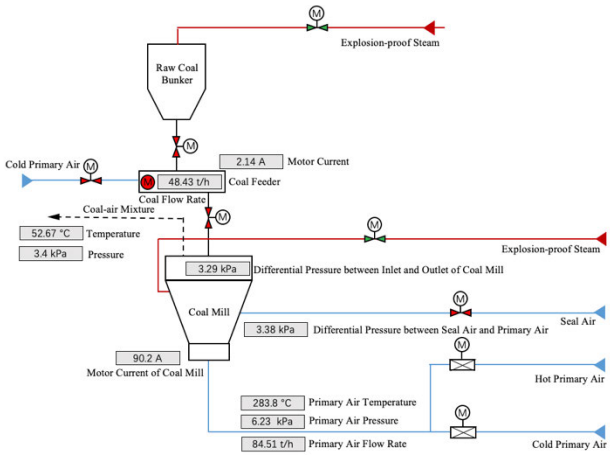


FIGURE 1. Structure of the medium speed coal pulverizing system.

B. COMMON FAULTS

During the operation process, common faults such as foreign materials intrusion, coal choking, coal shortage, mill fire or explosion, and etc. may occur [8].

1) Foreign materials intrusion: In the process of mining, transportation and storage, the raw coal is easily mixed with various types of foreign objects such as stones. Once some hard-to-grind objects with a considerable size are sent to the coal mill system, it may result in the jamming of the coal

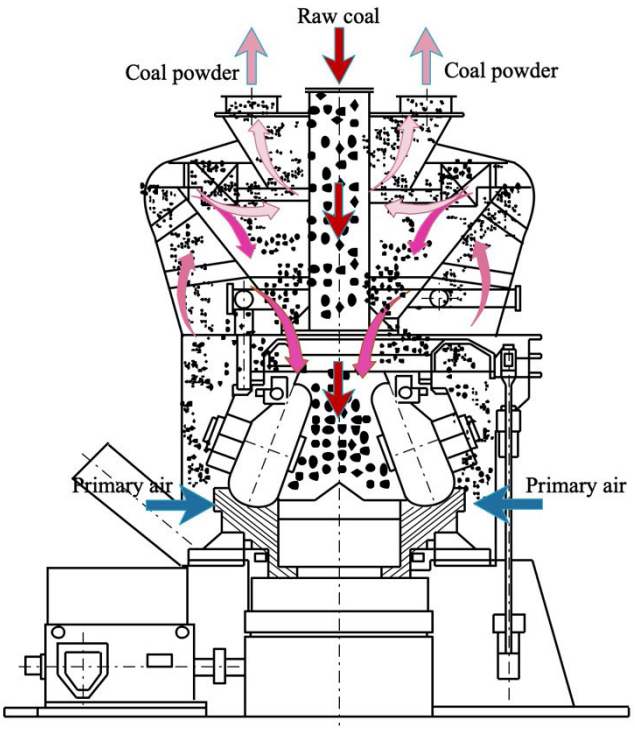


FIGURE 2. Schematic diagram of ZGM-113N medium speed coal mills.

falling pipe, severe vibration or even mechanical failure of the coal mill.

2) Coal choking: If the chocking occurs, the flow rate of coal-air mixture for burning will decrease, and the amount of accumulated coal inside the mill will keep increasing. The possible reasons are huge changes in coal quality, high moisture content in the raw coal, improper adjustment of the air-to-coal ratio, too small primary air flow rate, etc. The specific symptoms include the increase of differential pressure between the inlet and the outlet of the coal mill, the increase of the mill current, and the decrease of outlet temperature.

3) Coal shortage: When such fault happens, the amount of coal inside the mill will decrease. The common reasons are blockage of the coal falling pipe, malfunctions of the coal feeder, etc. The phenomena of such fault are the decrease of differential pressure between the inlet and the outlet of the coal mill, the reduction of the mill current, and the increase of the outlet temperature.

4) Mill fire or explosion: If this fault occurs, it is manifested by a sharp increase in the outlet temperature, a significant change in the wind pressure at the mill inlet, and even the combustion or ignition of coal inside the mill. The main causes are high primary air temperature, low coal flow rate, coal accumulation in the mill, coal leakage, etc.

Whatever the kind of fault happens, the later such fault is discovered, the greater the irreversible damage is caused to. Therefore, timely discovery and maintenance of these faults do help improve safety and reduce economic losses.

III. PRELIMINARIES

A. GAUSSIAN MIXTURE MODEL

A GMM combines the advantages of both parametric and non-parametric estimation, which is not limited to a specific function form. GMM can provide superior capability of dividing a sample dataset into different clusters by local Gaussian distribution, and each Gaussian component corresponds to one possible operating mode.

Let $\boldsymbol{x} = [x_1, x_2 \ldots x_m]^{\mathrm{T}}$ be a m-dimensional vector collected from coal mills, its probability density $p(\boldsymbol{x}|\boldsymbol{\Theta})$ can be expressed by different Gaussian components:

$$p(\boldsymbol{x}|\boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k N(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

where $\boldsymbol{\Theta} = \{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, (\boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K), \pi_1, \ldots, \pi_K\}$, $\boldsymbol{\mu}_k$ is the k^{th} mean vector, $\boldsymbol{\Sigma}_k$ is k^{th} covariance matrix, π_k is the prior probability of the k^{th} Gaussian component and satisfies $\sum_{k=1}^{K} \pi_k = 1$. K represents the number of total Gaussian components, and the function $N(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the multivariate Gaussian density function given by:

$$N\left(\boldsymbol{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) = \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}_k\right)^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}_k\right)\right\} \tag{2}$$

It is noted that given a number of n independent and identically distributed data $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_n]^T \in \mathbb{R}^{n \times m}$, the observed data log-likelihood function is given by

$$\log p(\mathbf{X} \mid \boldsymbol{\Theta}) = \log \prod_{i=1}^{n} p(\boldsymbol{x}_i \mid \boldsymbol{\Theta}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \pi_k N(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{3}$$

In general, parameters can be obtained by computing the maximum likelihood estimation (MLE) as follows:

$$\boldsymbol{\Theta}_{\mathrm{ML}} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} \{\log p(\mathbf{X}|\boldsymbol{\Theta})\} \tag{4}$$

The expectation-maximization (EM) algorithm gets around this issue by applying two steps iteratively until meeting the convergence criterion. The hyperparameters $\boldsymbol{\Theta}$ can be estimated by the EM algorithm. More details can be found in ref. [41], [51].

B. MULTI-OUTPUT RELEVANCE VECTOR MACHINE

Given a set of input variables $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots \boldsymbol{x}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times m}$ and target outputs $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2 \ldots \boldsymbol{y}_n]^{\mathrm{T}} \in \mathbb{R}^{n \times v}$, the nonlinear relationship between inputs and targets can be expressed as

$$\mathbf{Y} = \langle \boldsymbol{\Phi}, \mathbf{W} \rangle + \mathbf{E} \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product, $\boldsymbol{\Phi} = [\boldsymbol{\phi}(\boldsymbol{x_1})\, \boldsymbol{\phi}(\boldsymbol{x_2}) \ldots \boldsymbol{\phi}(\boldsymbol{x_n})]^{\mathrm{T}}$ is a design matrix with $\boldsymbol{\phi}(\boldsymbol{x}) = [1k(\boldsymbol{x}, \boldsymbol{x_1})\, k(\boldsymbol{x}, \boldsymbol{x_2}) \ldots k(\boldsymbol{x}, \boldsymbol{x_n})]^{\mathrm{T}}$, $k(\boldsymbol{x}, \boldsymbol{x_i})$ is the kernel function. $\mathbf{W}$ is the weight matrix, $\mathbf{E} = [\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_2 \ldots \boldsymbol{\varepsilon}_n]^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}_i$ are independent samples obeying the Gaussian distribution with zero mean and a covariance matrix $\boldsymbol{\Omega} = (\mathbf{E}^{\mathrm{T}}\mathbf{E})/n \in \mathbb{R}^{v \times v}$.

With the assumption that $\boldsymbol{\varepsilon}_i$ follow an independent Gaussian distribution, the likelihood of the training data can be expressed as

$$p(\mathbf{Y} \mid \mathbf{W}, \Omega) = (2\pi)^{-\frac{vn}{2}} |\boldsymbol{\Omega}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}(\mathbf{Y} - \Phi\mathbf{W})^{\mathrm{T}}(\mathbf{Y} - \Phi\mathbf{W}))\right) \tag{6}$$

where $\mathrm{tr}(\cdot)$ denotes trace of matrix.

To overcome the over-fitting problem, the estimation of weight matrix $\mathbf{W}$ is introduced by

$$p(\mathbf{W} \mid \boldsymbol{\alpha}, \boldsymbol{\Omega}) = (2\pi)^{-\frac{v(n+1)}{2}} |\boldsymbol{\Omega}|^{-\frac{n+1}{2}} |\mathbf{A}|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{W})\right) \tag{7}$$

where $\mathbf{A}^{-1} = \mathrm{diag}\left(\alpha_0^{-1}\alpha_1^{-1} \ldots \alpha_n^{-1}\right) = \frac{\mathbb{E}(\mathbf{w}\mathbf{w}^{\mathrm{T}})}{\mathrm{tr}(\boldsymbol{\Omega})}$, with $\mathbf{w}$ being the column vectors of $\mathbf{W}$. $\mathbf{W}$ keeps to the zero-mean Gaussian distribution with inverse variances $\boldsymbol{\alpha} = [\alpha_0 \alpha_1 \ldots \alpha_n]^{\mathrm{T}}$.

According to the Bayes' rules, the posterior probability over matrix $\mathbf{W}$ can be represented as follows:

$$p(\mathbf{W} \mid \mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) = \frac{p(\mathbf{Y} \mid \mathbf{W}, \boldsymbol{\Omega})\, p(\mathbf{W} \mid \boldsymbol{\alpha}, \boldsymbol{\Omega})}{p(\mathbf{Y} \mid \boldsymbol{\alpha}, \boldsymbol{\Omega})} \tag{8}$$

It can be further deduced to the following formulation:

$$p(\mathbf{W} \mid \mathbf{Y}, \boldsymbol{\alpha}, \boldsymbol{\Omega}) = (2\pi)^{-\frac{v(n+1)}{2}} |\boldsymbol{\Omega}|^{-\frac{n+1}{2}} |\boldsymbol{\Sigma}|^{-\frac{v}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Omega}^{-1}(\mathbf{W} - \mathbf{M})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{W} - \mathbf{M}))\right) \tag{9}$$

where $\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \mathbf{A})^{-1}$ and $\mathbf{M} = \boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y}$.

Great details regarding the mathematical theorem and the solving process are provided in ref[35], [36]. During the process of training MRVR, the application of EM algorithm helps obtain an optimal set of parameters. When handling multi-output regression problems, the optimal weight matrix $\mathbf{M}$ can be calculated with the obtained parameters.

C. BAYESIAN MULTI-OUTPUT RELEVANCE VECTOR RECONSTRUCTION

As aforementioned, to achieve great map relationships between inputs and reconstructions with MRVR, all faulty data of coal mills are needed for training. Considering the limitations of applying MRVR in the PM-FD of coal mills, a reconstruction method based on Bayesian inference is designed, namely Bayesian multi-output relevance vector reconstruction (BMRVR). The highlight of BMRVR is that a probabilistic mapping between inputs and targets is established, and the requirement for faulty training data is eliminated.

In line with MRVR, a reconstruction vector $\boldsymbol{x}^*$ can be calculated from a new input vector $\boldsymbol{x}$, which is provided by

$$\boldsymbol{x}^* = \mathbf{M}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}) \tag{10}$$

with predictive covariance matrix $\boldsymbol{\Omega}_*$ can be expressed as

$$\boldsymbol{\Omega}_* = \boldsymbol{\Omega}_{\mathrm{MP}} + \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}(\boldsymbol{x}) \tag{11}$$

where $\boldsymbol{\Omega}_{\mathrm{MP}}$ represents the estimated noise of the training data under normal conditions. As proposed in ref. [33,36]

the predictive covariance matrix $\boldsymbol{\Omega}_*$ is a sum of two terms, among them $\boldsymbol{\Omega}_{\mathrm{MP}}$ is the noise on the training data and $\boldsymbol{\Omega}_{\mathrm{MP}}\boldsymbol{\phi}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\phi}(\boldsymbol{x})$ represents the uncertainty in the prediction of the weights.

In this framework it is significant to note that (10) can successfully predict the $\boldsymbol{x}^*$ if the new $\boldsymbol{x}$ is normal, otherwise it is hard to reconstruct the fault data to normal state because of lacking all possible deviations when training. To overcome this problem, a Gaussian random vector $\boldsymbol{\varepsilon}^*$ with zero mean and an unknown diagonal covariance matrix $\boldsymbol{\Xi}$ is introduced. Then the (10) can be re-expressed as

$$\boldsymbol{x}^* = \mathbf{M}^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{x}) + \boldsymbol{\varepsilon}^* \tag{12}$$

Hence the following step is to estimate the deviations $\boldsymbol{\varepsilon}^*$ through $\boldsymbol{\Xi}$. Since the parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ of the k^{th} Gaussian component are available, the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{x}^*$ can be analytically computed via

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x}^* \end{bmatrix} | k \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_k \\ \boldsymbol{\mu}_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_k & \boldsymbol{\Sigma}_k \\ \boldsymbol{\Sigma}_k & \boldsymbol{\Sigma}_k + \boldsymbol{\Xi} \end{bmatrix}\right) \tag{13}$$

It is not surprising to resort to EM for the estimation of $\boldsymbol{x}^*$ and $\boldsymbol{\Xi}$. In this case, the purpose of EM is to solve the following maximum likelihood estimation problem:

$$\hat{\boldsymbol{\Xi}}^* = \underset{\boldsymbol{\Xi}}{\operatorname{argmax}}\, p(\boldsymbol{x}|\boldsymbol{\Xi}) \tag{14}$$

Let $L(\boldsymbol{\Xi}) = \log p(\boldsymbol{x}|\boldsymbol{\Xi})$ be the observed data log likelihood, after l iterations, the optimal estimated parameter is $\hat{\Xi}_l^*(l \in \mathbb{N})$. The quantity of change has the sign as

$$L(\boldsymbol{\Xi}) - L\left(\hat{\boldsymbol{\Xi}}_l^*\right) = \log \frac{\sum_{\boldsymbol{x}}^* p(\boldsymbol{x}|\boldsymbol{x}^*, \boldsymbol{\Xi}) p(\boldsymbol{x}^*|\boldsymbol{\Xi})}{p\left(\boldsymbol{x} \mid \hat{\boldsymbol{\Xi}}_l^*\right)} \frac{p(\boldsymbol{x}^*|\boldsymbol{x}, \hat{\boldsymbol{\Xi}}_l^*)}{p(\boldsymbol{x}^*|\boldsymbol{x}, \hat{\boldsymbol{\Xi}}_l^*)} \tag{15}$$

In light of the Jensen inequality, the change of likelihood given by (15) is modified to

$$L(\boldsymbol{\Xi}) \geq L\left(\hat{\boldsymbol{\Xi}}_l^*\right) + \Delta(\boldsymbol{\Xi}|\hat{\boldsymbol{\Xi}}_l^*) \tag{16}$$

where

$$\Delta(\boldsymbol{\Xi}|\hat{\boldsymbol{\Xi}}_l^*) = \sum\nolimits_{\boldsymbol{x}^*} p(\boldsymbol{x}^*|\boldsymbol{x}, \hat{\boldsymbol{\Xi}}_l^*) \log \frac{p(\boldsymbol{x}|\boldsymbol{x}^*, \boldsymbol{\Xi}) p(\boldsymbol{x}^*|\boldsymbol{\Xi})}{p(\boldsymbol{x}|\hat{\boldsymbol{\Xi}}_l^*) p(\boldsymbol{x}^*|\boldsymbol{x}, \hat{\boldsymbol{\Xi}}_l^*)}.$$

To investigate further, the aim of the E-step is to compute the following equation:

$$Q(\boldsymbol{\Xi}|\hat{\boldsymbol{\Xi}}_l^*) = E_{\boldsymbol{x}^*|\boldsymbol{x}, \hat{\boldsymbol{\Xi}}_l^*} \log p(\boldsymbol{x}, \boldsymbol{x}^*|\boldsymbol{\Xi}) \tag{17}$$

It is natural to seek $\hat{\boldsymbol{\Xi}}_{l+1}^*$ in the M-step via

$$\hat{\boldsymbol{\Xi}}_{l+1}^* = \underset{\boldsymbol{\Xi}}{\operatorname{argmax}}[Q(\boldsymbol{\Xi}|\hat{\boldsymbol{\Xi}}_l^*)] \tag{18}$$

When the difference between $\hat{\boldsymbol{\Xi}}_{l+1}^*$ and $\hat{\boldsymbol{\Xi}}_l^*$ is practically equal to zero, it means that the reconstruction process has reached convergence and the observed input has been adjusted to a formal state.

IV. PROPOSED MULTI-MODE FAULT DIAGNOSIS METHOD

The proposed algorithm makes use of GMM and BMRVR to design a novel combination strategy for PM-FD of coal mills corresponding to different running modes. As shown in Fig. 3, the entire framework of the multi-mode fault diagnosis method consists of two main parts, namely offline training and online application. The detailed implementation steps are given in Algorithm A as well.

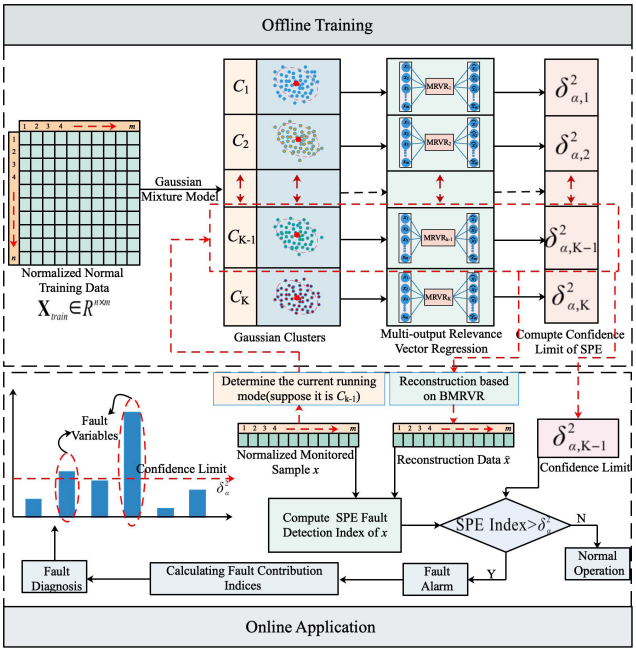


FIGURE 3. Schematic diagram of the proposed fault diagnosis method.

A. OFFLINE TRAINING

After completing the normalization procedure of the collected normal operating data of coal mills, GMM can be established as the first step to divide the training samples into several running modes. The prior probability of the kth Gaussian component can be given by π_k, and the probability density can be calculated by the weighted sum of different Gaussian components according to (1). After obtaining the GMM model, the local MRVR algorithm is taken for training the initial relationships of different Gaussian batches.

As described in the above part, the GMM and MRVR models are successfully obtained. However, a complete fault diagnosis algorithm should identify the normal or abnormal status of monitored samples correctly. As a result, it is essential to compute the control limits satisfying the requirement of a certain confidence. In this paper, SPE [41,50] is chosen as the fault detection index, which is defined as the squared norm of the difference vector between the measurement $\boldsymbol{x}$ and the reconstructed $\boldsymbol{x}^*$. SPE can be calculated as:

$$\mathrm{SPE} = \left\|\boldsymbol{x} - \boldsymbol{x}^*\right\|^2 = \sum_{i=1}^{m} (x_i - x_i^*)^2 \tag{19}$$

with the control limit [25] is given by

$$\delta_\alpha^2 = \theta_1 \left(1 - \frac{\theta_2 h_0 (1-h_0)}{\theta_1^2} + \frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} \right)^{\frac{1}{h_0}} \tag{20}$$

where z_α is denoted as the normal variable corresponding to $(1-\alpha)\times 100$th percentile, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, $\theta_i = \sum_{k=1}^{m} \lambda_k^i$, and λ_k^i are the eigenvalues of the sample covariance matrix $\mathbf{S}$. $\mathbf{S}$ can be calculated from the residual matrix $\mathbf{E} = (\mathbf{X} - \mathbf{X}^*)$, which takes the following formulation:

$$S(j,k) = \frac{1}{n} \sum_{i=1}^{n} e_{ij} e_{ik} \tag{21}$$

where $e_{ij} = \mathbf{E}(i,j)$ is the $(i,j)^{\text{th}}$ value of the residual matrix $\mathbf{E}$.

The sample is considered normal if its SPE is within the control limit δ_α^2, otherwise it is treated in a faulty state.

B. ONLINE APPLICATION

When a new monitored sample $\boldsymbol{x}$ comes, it should first be normalized the same way as the training samples does. The second step is to compute the posterior probabilities and identify the Gaussian component to which the vector belongs. Thirdly, the reconstruction data $\boldsymbol{x}^*$ is calculated with the application of the local BMRVR.

The objective of the practical application is to determine whether the testing data is under normal operation or faulty condition. As a result, the main index SPE which depends on the monitored sample and reconstruction data is further calculated. If the SPE is smaller than δ_α^2, the given sample is defined as normal data, otherwise it is defined as fault data. It should be noted that we need to perform root cause analysis when a sample is identified as faulty. Contribution plot is an effective method, and a variable is most likely the faulty one if it has the largest contribution to SPE statistics. The expressions to compute all contribution indices of the multivariate samples have the quadratic form, which can be written as:

$$C_i^{\text{SPE}} = (x_i - x_i^*)^2 \tag{22}$$

The above contribution index indicates the magnitude of the possibility that a specific variable has the faulty status. In other words, the variables are more likely responsible for the abnormality if the contribution indices are significantly larger than those of other variables.

To evaluate the estimated results of our reconstruction method given the Gaussian assumption of residuals, the root means square error (RMSE) and prediction efficiency (PE as a percentage) [52] are added and analyzed in this paper. RMSE denotes the actual reconstruction deviation and PE reflects the relative level of deviation. When the RMSE is small, or the PE is close to 100%, the deviations indicates an excellent reconstruction method. RMSE and PE are calculated as

$$\text{RMSE}_i = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(x_{i,j}^* - x_{i,j} \right)^2} \tag{23}$$

Algorithm 1 The proposed fault diagnosis algorithm integrating GMM and BMRVR

Inputs:
Multivariate normal training data matrix $\mathbf{X}_{\text{train}} \in \mathbb{R}^{n\times m}$
Require:
Initial parameters
$K_{\min}, K_{\max}, \boldsymbol{\Theta} = \left\{ (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, \left(\boldsymbol{\mu}_{K_{\max}}, \boldsymbol{\Sigma}_{K_{\max}}\right), \pi_1, \ldots, \pi_{K_{\max}} \right\}$
(Offline training)
$\boldsymbol{s} \leftarrow 0, \boldsymbol{K} \leftarrow \boldsymbol{K_{max}}$
while (stopping criterion is not met) **do**
  Calculate $p^{(s)}(\mathbf{x}_i \in C_k)$ (**E-step**)
  Calculate $\boldsymbol{\mu}_k^{(s+1)}, \boldsymbol{\Sigma}_k^{(s+1)}, \pi_k^{(s+1)}$ (**M-step**)
  Update parameters $\boldsymbol{\Theta}_{\text{ML}} \leftarrow \underset{\boldsymbol{\Theta}}{arg\ \max} \{\log p(\mathbf{X}|\Theta)\}$
  $K \leftarrow K - 1$, if $\pi_k^{(s+1)} \le \epsilon$ (ϵ is a small positive number)
  $s \leftarrow s + 1$
end while
Divide training data $\mathrm{X}_{\text{train}}$ into K Gaussian clusters with GMM
for $i = 1$ to K **do**
  Estimate hypermeters $\boldsymbol{\alpha}_i$, $\boldsymbol{\Sigma}_i$, $\mathbf{M}_i$ and $\boldsymbol{\Omega}_i$ of the MRVR model
  Calculate control limits $\delta_{\alpha,i}^2$ of all gaussian clusters via (20)
end for
(Online application)
for $j = 1$ to n_{test} **do**
  Determine the current running-on mode of the monitored $\boldsymbol{x}_j$ with GMM, and pick up the corresponding confidence limits $\delta_{\alpha,i}^2$
  Calculate the expectation with local MRVR using (10)
  Repeat the E-step via (17) and M-step using (18) until the algorithm converges and output the final reconstruction data $\boldsymbol{x}_j^*$, namely BMRVR.
  Calculate the squared prediction error (SPE) via (19)
  if SPE $< \delta_{\alpha,i}^2$ **then**
    Declare the monitored sample is in a normal state
  else
    Declare fault of the monitored sample and perform root cause analysis
  end if
end for

$$\text{PE}_i(\%) = \left(1 - \frac{\sum_{j=1}^{n} \left(x_{i,j}^* - x_{i,j} \right)^2}{\sum_{j=1}^{n} \left(x_{i,j}^* - \bar{x}_i \right)^2}\right) \times 100 \tag{24}$$

where $i = (1, 2, \cdots, m)$ represents i^{th} variable from the m monitored variables, and $\bar{x}_i$ denotes the actual mean value of i^{th} variable.

In addition, two statistics, namely fault detection rate (FDR) and false alarm rate (FAR), are used to access the effectiveness of the fault diagnosis performance. FDR and

FAR are defined and depicted by:

$$\mathrm{FDR} = \frac{TP}{TP + FN} \tag{25}$$

$$\mathrm{FAR} = \frac{FP}{FP + TN} \tag{26}$$

where TP is the number of isolated abnormal variables with faulty label, FN is the number of faulty variables determined to be normal, FP is the number of variables that are incorrectly judged to be abnormal under the normal state, and TN is the number of normal variables that are not falsely detected.

What is more, the computing time (CT) for each sample is another index related to the computational efficiency of different PM-FD methods.

V. CASE STUDY

In this section, the proposed multi-mode PM-FD method is evaluated using a self-defined nonlinear system and two real fault cases of a medium speed coal mill. Following the data collection and preprocessing, the procedure of the proposed PM-FD scheme proceeds in five main steps:

Step 1: Establish GMM model with the training samples under normal conditions, which estimates parameters $\boldsymbol{\Theta} = \{(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}), \ldots, (\boldsymbol{\mu_K}, \boldsymbol{\Sigma_K}), \boldsymbol{\pi_1}, \ldots, \boldsymbol{\pi_K}\}$ according to equations (1)-(4).

Step 2: Divide training samples into $\boldsymbol{K}$ Gaussian clusters with the established GMM model.

Step 3: For all Gaussian clusters, establish local MRVR models one by one, estimate hypermeters $\boldsymbol{\alpha_i}$, $\boldsymbol{\Sigma_i}$, $\mathbf{M_i}$ and $\boldsymbol{\Omega_i}$ using (6)-(9), and compute confidence limits $\boldsymbol{\delta}^2_{\boldsymbol{\alpha},\boldsymbol{i}}$ ($\boldsymbol{i} = 1, 2, \cdots, \boldsymbol{K}$) in (20).

Step 4: Determine the current running-on mode of newly observed sample with GMM, calculate the expectation with corresponding local MRVR according to (10), and the expectation is further reconstructed with the application of the Bayesian improved algorithm (BMRVR) using (12).

Step 5: Calculate the fault index SPE via (19) and, finally, fault detection and diagnosis by comparing fault indices with confidence limits.

To further testify the superiority of the proposed method, different approaches e.g. PCA [22], KPCA [23], KPLS [26], AANN [28], MRVR, multivariate Gaussian process regression (MGPR) [50] and their GMM-based variants are used to make comparisons. For these GMM-based variants, GMM are first applied for data classification, and the corresponding local PM-FD methods are further utilized for fault detection.

The configuration of the computational platform is specified as follow: 1) the CPU is an Intel processer, E5-2643 3.4GHz; 2) the RAM volume is 16.00G; 3) the Operation system is Windows 10; and 4) the coding environment is Python v3.7.

A. VALIDATION BY A SELF-DEFINED NONLINEAR SYSTEM

To illustrate the validity of the proposed fault diagnosis method, a simulation study is carried out in this section. We first introduce a strong nonlinear simulated process with six variables. The process established by self-defined functions is given by

$$\begin{cases} x_1 = u_1 + \gamma_1 \\ x_2 = 0.2(u_1^2 - 2u_1 + 3) + \gamma_2 \\ x_3 = 2e^{u_1/10} + \gamma_3 \\ x_4 = 0.8u_2 + 1.5 + \gamma_4 \\ x_5 = \cos(u_2 + \gamma_5 \, x_6 = 1.1|u_1 - u_2| + \gamma_6 \end{cases} \tag{27}$$

where $\gamma_i \sim N(0, 0.05)$ is the normally distributed noise. Three different operating modes are studied to analyze the performance of the proposed multi-mode, which can be illustrated as: mode 1: $u_1 \sim N(4, 0.25)$, $u_2 \sim N(12, 0.8)$, mode 2: $u_1 \sim N(7, 0.45)$, $u_2 \sim N(2, 0.15)$, mode 3: $u_1 \sim N(11, 0.65)$, $u_2 \sim N(15, 0.7)$. A total of 2400 normal operating samples (with each mode 800) are generated according to (27).

All samples are first scaled to the range of [0,1] via the following equation:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{28}$$

where x and x' represent the original and scaled variable, respectively. $x_{\max}$ and $x_{\min}$ denote the maximum and minimum value of the raw data. Another set of 2400 (with each mode 800) testing samples are obtained according to (27) as well. As described in Table 1, four different types of faults (200 samples with each) are introduced to every mode.

TABLE 1. Faults status and corresponding description.

Fault status	Description
Fault0	Normal condition without fault
Fault1	A step fault with the magnitude of 3 is introduced to x_2
Fault2	A step fault with the magnitude of 3 is introduced to x_6
Fault3	A step fault with the magnitude of 1 is introduced to x_1 and 1.5 is introduced to x_4

The experimental test on the self-defined system is carried out in this part. As shown in Fig. 4, the performance of operation mode identification is sufficient to meet the requirements for both training and testing data. Fig. 5 illustrates the detection results of the three different fault statuses for mode1, mode2 and mode3, respectively. As expected, the proposed algorithm can capture data anomalies once a fault occurs.

After applying the proposed reconstruction model, we calculate the RMSEs and PEs of all variables under different modes and fault types. As shown in TABLE 2, RMSEs of all variables are quite small and the corresponding PEs are relatively high. It is apparent that the reconstruction capacity of the proposed GMM-BMRVR approach is effective.

The experimental test on the self-defined system is also carried out in this part. As shown in Fig. 4, the performance of operation mode identification is sufficient to meet the requirements for both training and testing data. Fig. 5 illustrates the detection results of the three different fault statuses

TABLE 2. Reconstruction performances of the proposed BMRVR under different modes and faults.

Mode		x_1		x_2		x_3		x_4		x_5		x_6	
		RMSE	PE	RMSE	PE	RMSE	PE	RMSE	PE	RMSE	PE	RMSE	PE
Mode1	Fault0	0.23	95.43	0.35	93.23	0.12	94.81	0.51	92.65	0.46	91.36	0.53	90.12
	Fault1	0.26	94.21	0.47	88.76	0.08	96.58	0.53	90.31	0.41	96.25	0.85	87.08
	Fault2	0.24	94.53	0.29	94.35	0.31	89.79	0.62	82.71	0.62	79.48	1.06	81.22
	Fault3	0.49	86.06	0.27	96.24	0.08	93.13	0.97	80.31	0.39	86.33	0.96	82.05
Mode2	Fault0	0.33	93.74	0.41	94.31	0.22	94.32	0.26	93.29	0.2	94.61	0.69	88.52
	Fault1	0.41	89.82	0.95	80.12	0.2	95.33	0.13	96.47	0.15	95.49	0.55	89.36
	Fault2	0.44	90.23	0.51	90.67	0.18	94.46	0.12	95.12	0.14	95.08	0.98	79.35
	Fault3	0.58	77.46	0.42	89.53	0.21	90.27	0.76	82.21	0.35	88.22	0.56	89.05
Mode3	Fault0	0.36	90.68	0.38	91.11	0.37	90.45	0.5	91.59	0.39	87.51	0.77	88.75
	Fault1	0.39	91.72	0.66	87.45	0.38	89.48	0.57	85.25	0.63	81.26	0.96	83.56
	Fault2	0.33	90.78	0.37	93.22	0.35	90.62	0.53	88.36	0.71	80.13	0.92	82.35
	Fault3	0.71	70.43	0.74	83.45	0.46	87.92	0.67	82.16	0.43	84.38	1.03	80.56

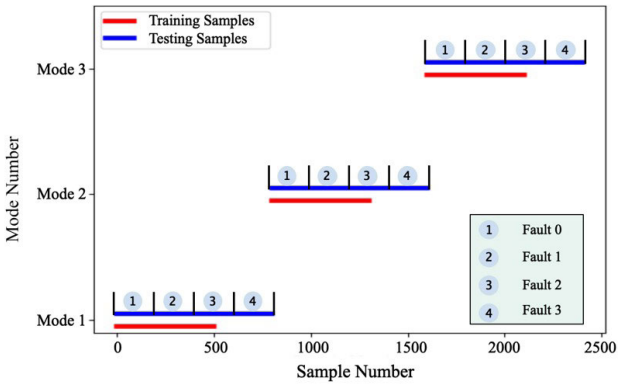


FIGURE 4. Classification results of different operation modes for both training and testing samples.

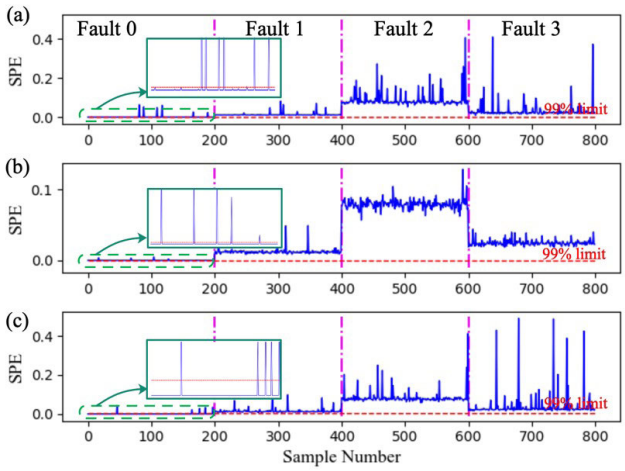


FIGURE 5. Training results of three different fault statuses for (a) mode1, (b)mode2, (c)mode3.

for mode1, mode2 and mode3, respectively. As expected, the proposed algorithm can capture data anomalies once a fault occurs.

To further isolate the faults, the contribution plot of three different fault statuses for mode1, 2 and 3 is shown in Fig. 6. As depicted, under the condition of fault 1, the other five SPE contributions are smaller than the control limit. When fault 2 and 3 are introduced to testing samples, although the

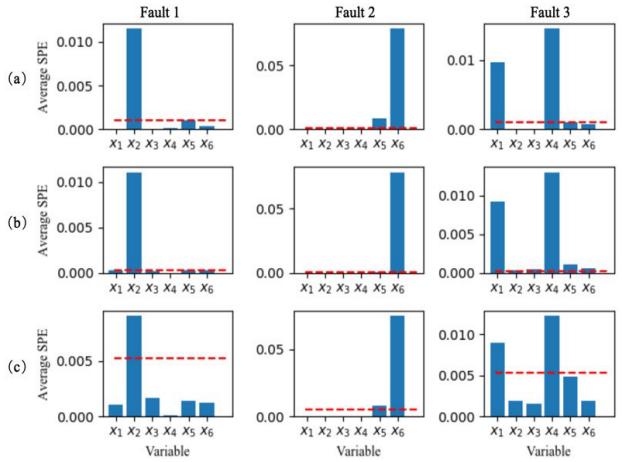


FIGURE 6. Training results of three different fault statuses for (a) mode1, (b)mode2, (c)mode3.

contribution of a specific variable slightly exceeds the 99% control limit, the contributions of fault variables are much higher than the normal ones, and the effect of isolating faults is significant.

To demonstrate the advantages of the proposed GMM-BMRVR method, a performance comparison of different approaches is carried out. The average evaluated indices (FDR, FAR) as well as CT are listed in TABLE 3. It is apparently shown that the FDR values corresponding to the three different types of faults are all over 95%. Correspondingly, FARs fluctuate in a small range of around 5%. For a piece of testing sample, the average CT of 2.15 milliseconds is reasonable as well. It is noticed from Table 3 that GMM-based methods provide better results compared to those classical single PM-FD methods like PCA [22], AANN [28] and KPLS [26]. On the other hand, for GMM-like methods, when GMM-PCA and GMM-AANN are utilized to deal with the fault data, the corresponding FDRs are smaller and FARs fluctuate significantly. In particular, GMM-PCA is inadequate when facing the complex nonlinear systems. GMM-KPCA, GMM-KPLS, GMM-MGPR and GMM-MRVR produce relatively good results when handling the given nonlinear system, but there is still some gap compared with GMM-BMRVR.

TABLE 3. Comparison of different algorithms in the case of isolating faults.

Model	Fault1		Fault2		Fault3		
	FDR (%)	FAR (%)	FDR (%)	FAR (%)	FDR (%)	FAR (%)	CT (ms)
PCA[22]	68.9	30.23	84.6	14.54	72.3	19.34	0.89
AANN[28]	80.4	20.48	82.8	10.79	89.6	10.76	2.23
KPCA[23]	82.3	12.34	90.5	11.06	89.3	12.48	3.02
KPLS[26]	81.4	12.39	83.5	12.43	86.2	13.65	2.55
MRVR	82.1	14.82	87.5	12.67	87.2	14.56	2.85
MGPR[50]	83.7	14.39	88.7	10.47	90.3	12.13	3.38
BMRVR	87.4	12.45	91.2	8.34	88.6	13.98	3.46
GMM-PCA[44]	70.3	22.48	91.7	10.83	79.6	15.75	0.88
GMM-AANN	83.3	18.37	87.9	8.83	86.8	20.43	1.76
GMM-KPCA	91.3	4.74	100	3.12	92.7	8.91	2.31
GMM-KPLS	91.8	7.36	99.5	6.79	91.4	7.65	2.02
GMM-MRVR	92.6	6.74	100	5.87	92.8	6.83	1.90
GMM-MGPR	92.1	5.28	100	5.32	92.5	6.92	2.69
GMM-BMRVR	**95.6**	**4.25**	**100**	**3.63**	**95.4**	**5.32**	**2.15**

To sum up, the proposed GMM-BMRVR performs well in PM-FD problems of different fault types.

B. FAULT DIAGNOSIS OF A MEDIUM SPEED COAL MILL

The work in this part is based on a ZGM-113N medium speed coal mill equipped in a subcritical 600MW unit. Combined with existing research [1], [53] and relevant theoretical knowledge [54], 15 operating variables listed in Table 4 are selected to establish a coal mill fault diagnosis model. The coal mill studied in this paper was thoroughly overhauled, and, after that, a total of 28,800 historical data samples were acquired for ten days. The sampling interval is 30 seconds, and the range of unit load range is between 300 and 600MW. Due to the frequent changes of the grid load command, samples of different load segments are unevenly distributed. Therefore, to ensure the uniformity of the training samples, we divided the load range into 10 segments at an interval of 30MW and chose 400 samples in each load segment through random non-repeated sampling. Finally, 4000 pieces of data were selected as the training dataset to build the PM-FD model. In this section, two typical fault events are studied, namely mechanical failure and coal blockage.

The GMM model is firstly established based on the normal historical data acquired. All the historical data are divided into five running modes, and Fig. 7 shows the classification result of four main related variables, in which the size of scattered particles represents the coal flow rate from the coal feeder. After the completion of the GMM model, the local MRVR training models are established for each mode and the corresponding data cluster. Based on the above local MRVR models, the control limits for different modes are obtained.

In practical operation processes of medium coal mills, when a fault occurs, it is unrealistic to obtain the actual data under the corresponding normal running conditions. As a result, we tested the gmm-bmrvr method by adding step faults to two variables, namely coal-air mixture outlet temperature and the motor current of coal mill. To fully verify the performance of the proposed method, we varied the magnitude of the step faults from 20% to 100% with

TABLE 4. Operating variables for establishing fault diagnosis model.

No.	Variables	Description	Unit
M1	N_{unit}	Unit load	MW
M2	M_{coal}	Coal flow rate from coal feeder	t/h
M3	W_{air}	Primary air flow rate	t/h
M4	T_{air}	Primary air temperature	°C
M5	P_{air}	Primary air pressure	kPa
M6	I_{mill}	Motor current of coal mill	A
M7	I_{feed}	Motor current of coal feeder	A
M8	ΔP_{seal}	Differential pressure between seal air and primary air	kPa
M9	$T_{coal-air}$	Coal-air mixture outlet temperature	°C
M10	$P_{coal-air}$	Coal-air mixture outlet pressure	kPa
M11	ΔP_{in-out}	Differential pressure between inlet and outlet of coal mill	kPa
M12	T_{oil}	Lubricating oil temperature of coal mill	°C
M13	T_{tile}	Thrust tile temperature of coal mill	°C
M14	$T_{bearing}$	Motor bearing temperature of coal mill	°C
M15	T_{stator}	Motor stator winding temperature	°C

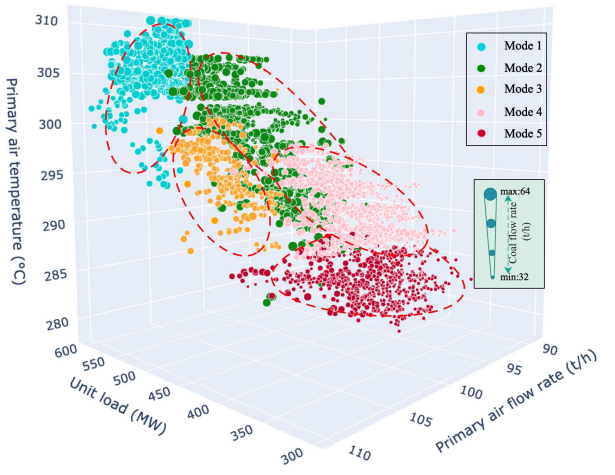


FIGURE 7. Operation modes classification result of the coal mill based on GMM.

the step size as 20. as shown in figure 8, the deviations between the reconstruction data and the actual normal values arE small for different fault magnitudes. RMSES and PES

TABLE 5. Reconstruction performances of two examples.

Variables	RMSE	PE
Coal-air mixture outlet temperature ($T_{coal-air}$)	0.89	92.05
Motor current of coal mill (I_{mill})	1.69	91.32

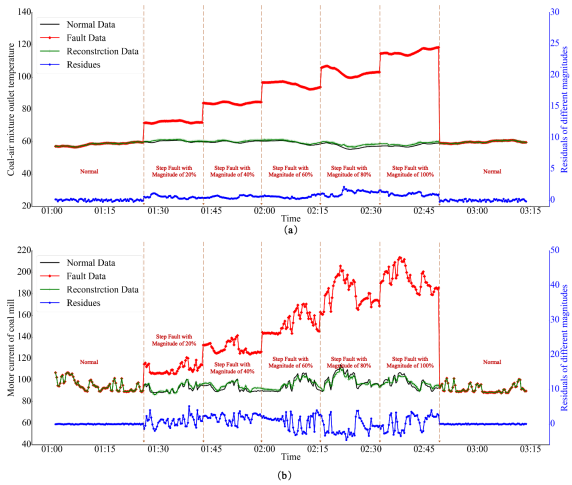


FIGURE 8. Examples of data reconstruction performance with the magnitude of step fault from 20% to 100%.

for those adding-faults data are also calculated after being reconstructed by the proposed method. As listed in Table 5, both rmses are small and the pes are higher than 90%, which confirms that the given method can correctly handle different faults. therefore, the gaussian assumption of residuals in our manuscript can satisfy the practical requirements.

1) FAULT CASE (1): MECHANICAL FAILURE

It started at 9:20 a.m., the coal flow rate oscillated around 40 t/h, but the current swung between 73-102A, which should have been less than 80A. Such phenomenon lasted for two hours before discovered. The fluctuation of coal mill current and the historical data plot of the coal flow rate are shown in Fig. 9. After the maintenance, it was found that there were some big foreign matters in the coal mill.

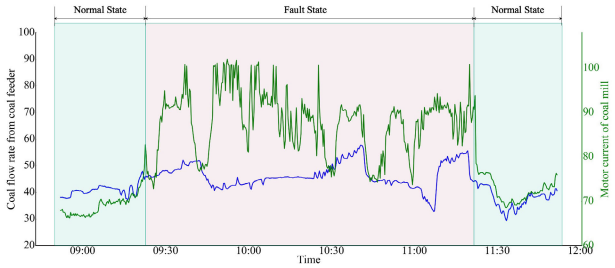


FIGURE 9. Historical variation trend of relevant parameters at different stages of fault case (1).

Here the testing results of fault case (1) are illustrated in Fig. 10 and Fig. 11. As discussed above, the primary

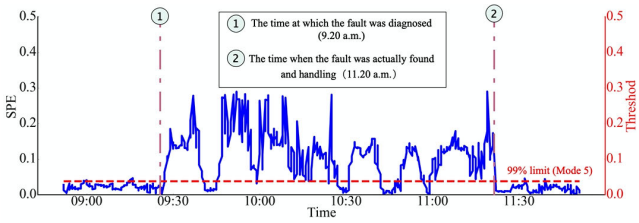


FIGURE 10. Testing result of the fault case (1) by the proposed method.

manifestation of the mechanical failure is the apparent fluctuation of coal mill current and the other parameters are relatively normal. As shown in Fig. 10, during the whole process, the coal mill's running mode is classified to mode 5. The SPE value begins to exceed the control limit from 9.30 a.m. and generally remains high. In particular, it is found that the value of SPE is lower than the 99% control limit in some periods. This is analytically due to the intermittent influence on the coal mill's operation when there is a foreign body, resulting in the motor current returning to normal operation at some moments. This situation will also lead to a reduction in FDR. However, as depicted in Fig. 11, the contribution plot of variables during the failure process reveals that the root lies in the current of coal mill. Moreover, the failure started at 9.20 a.m., but it was not discovered and dealt with by the operator until 11.20 a.m. As expected, the proposed fault diagnosis model significantly provides the early warning of the failure.

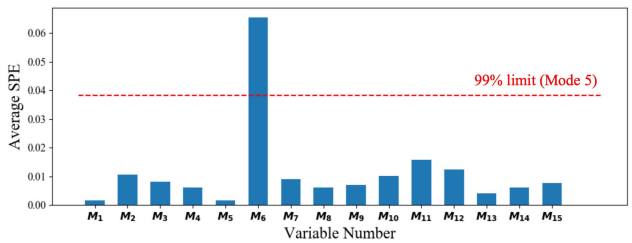


FIGURE 11. Contribution plot result of fault case (1).

2) FAULT CASE (2): COAL BLOCKAGE

As shown in Fig. 12, starting at 6.30 a.m., the current of the coal mill and the differential pressure between inlet and outlet began with oscillating rise. The upward trend became significant at 7.00 a.m. and continued until 7.40 a.m. From 6.30 a.m. to 7.40 a.m., the primary air flow rate was decreasing all the time. In addition, the coal-air mixture outlet temperature started to decline distinctly at 6.30 a.m. and remained low until 7.40 a.m. The coal mill began to block from 6.30 a.m., and it was not discovered and processed by the operator until 7.40 a.m., which seriously affected the normal operation of the equipment.

Figure 13 shows the testing result of the coal blockage by the proposed method. It is illustrated that the whole process went through three different operating modes, among which

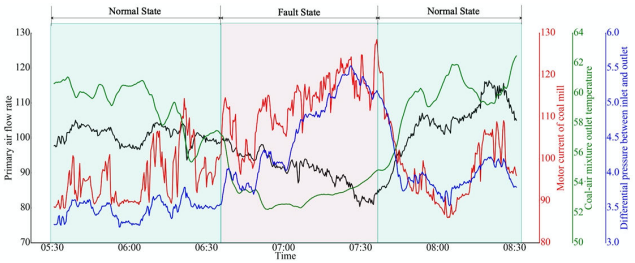


FIGURE 12. Historical variation trend of relevant parameters in different stages of fault case (2).

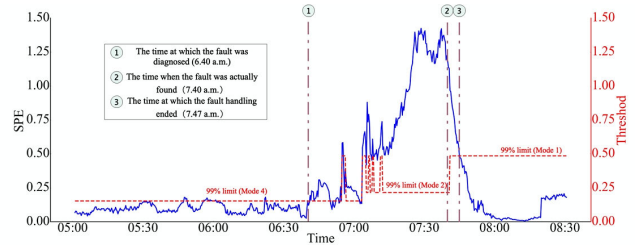


FIGURE 13. Testing result of the fault case (2) by the proposed method.

are mode 4, mode 2 and mode 1. The 99% control limits also vary with the operation mode. According to the on-site maintenance report, the fault started from 6.30 a.m. and was not discovered and handled until 7.40 a.m. With the application of the proposed method, the value of SPE begins to exceed the control limit at 6.40 a.m. and a significant deviation occurs at around 7.10 a.m. In other words, our method can give an alarm at least 30 minutes earlier than when the fault is actually found. Through the comprehensive analysis of Fig. 13 and Fig. 14(b), the period between 7.10 a.m. and 7.40 a.m. is the stage where the fault occurs significantly, with the corresponding operation mode is mode 2. During this period, the contribution values of the four main fault variables are evidential and the fault isolation is successfully reached. From 6.30 a.m. to 7.10 a.m., the failure is in the early stage and only the coal-air mixture outlet temperature changes

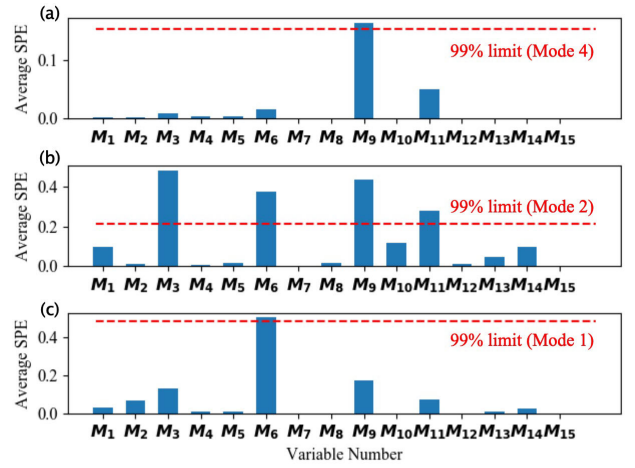


FIGURE 14. Contribution plot result of fault case (2) with (a) Mode4, (b) Mode 2, (c) Mode 1.

TABLE 6. Performance comparison results of different algorithms for isolating abnormal variables.

Model	Fault case (1)		Fault case (2)		CT
	FDR	FAR	FDR	FAR	(ms)
PCA[22]	81.5	14.56	65.4	20.3	2.35
AANN[28]	88.7	10.42	78.9	13.46	6.77
KPCA[23]	89.2	7.02	81.2	7.13	6.23
KPLS[26]	89.7	6.93	80.3	7.74	5.78
MRVR	88.5	7.19	81.9	8.98	7.10
MGPR[50]	90.1	6.23	82.8	7.29	7.73
BMRVR	91.5	5.43	84.6	4.87	7.84
GMM-PCA[44]	87.3	8.25	74.3	14.67	2.36
GMM-AANN	90.4	7.34	80.7	12.15	4.25
GMM-KPCA	91.1	3.12	86.4	3.73	4.91
GMM-KPLS	91.5	5.13	86.1	4.87	4.68
GMM-MRVR	91.3	4.57	85.7	6.32	4.12
GMM-MGPR	92.3	4.16	87.3	5.43	6.16
GMM-BMRVR	**94.1**	**2.65**	**89.4**	**2.13**	**4.84**

remarkably. Fig. 13(a) also shows that the contribution value of the temperature is higher than other variables. In addition, as shown in Fig. 13(c), the coal mill is in the recovery stage from 7.40 a.m., the corresponding contribution values gradually decrease and eventually fall below the control limit. It is worth mentioning that our model still has a small limitation. As revealed in Fig. 12, from 7.00 a.m. to 7.10 a.m., the misclassification problem caused by the proximity of the classification boundary limits the performance to some extent. However, despite such deficiencies, the performance of the model is still acceptable.

To further verify the superiority of the proposed GMM-BMRVR model, different fault diagnosis methods are utilized to make comparisons. TABLE 6 depicts the PM-FD results of single classic methods and GMM-based variants. In these two cases, all GMM-based methods like GMM-KPCA show their improvements at the level of both FDR and FAR to single methods like PCA. For those GMM-like methods, it turns out that GMM-BMRVR has larger FDR and smaller FAR for both fault case (1) and fault case (2), which means it performs better than others. Then come GMM-KPCA, GMM-MRVR, GMM-KPLS and GMM-MGPR, while the traditional GMM-PCA falls behind other models. The PCA-based model does not work very well when dealing with complex nonlinear processes such as coal mills. Moreover, in terms of CT, the proposed GMM-BMRVR has a satisfying performance.

In summary, after applying two fault cases in reality and comparing them with different methods, the proposed GMM-BMRVR model performs better.

VI. CONCLUSION

To tackle the PM-FD problem of complex non-linearity and multimodality of coal mills, a novel Bayesian multi-mode method is proposed in this paper. During the modeling

process, GMM is utilized for mode identification and MRVR is performed as a regression model to obtain the intrinsic knowledge and the control limit of the normal operating data in each mode. To overcome the problem of lacking all possible deviations in the training dataset, the Bayesian inference is adopted to help reconstruct the newly observed samples. The proposed method is evaluated by a numerical example and two realistic industrial fault cases. After the experimental verification, the proposed method is proved to have advantages in dealing with different kinds of faults of coal mills. In our further study, we will resort to reduce the redundant information of parameters and retain meaningful data to further enhance the model efficiency.

REFERENCES

[1] X. Hong, Z. Xu, and Z. Zhang, "Abnormal condition monitoring and diagnosis for coal mills based on support vector regression," *IEEE Access*, vol. 7, pp. 170488–170499, 2019, doi: 10.1109/ACCESS.2019.2955249.

[2] V. Agrawal, B. K. Panigrahi, and P. M. V. Subbarao, "Review of control and fault diagnosis methods applied to coal mills," *J. Process Control*, vol. 32, pp. 138–153, Aug. 2015, doi: 10.1016/j.jprocont.2015.04.006.

[3] S. Guo, J. Wang, J. Wei, and P. Zachariades, "A new model-based approach for power plant tube-ball mill condition monitoring and fault detection," *Energy Convers. Manage.*, vol. 80, pp. 10–19, Apr. 2014, doi: 10.1016/j.enconman.2013.12.046.

[4] J.-L. Wei, J. Wang, and Q. H. Wu, "Development of a multisegment coal mill model using an evolutionary computation technique," *IEEE Trans. Energy Convers.*, vol. 22, no. 3, pp. 718–727, Sep. 2007.

[5] Y. Hu, B. Ping, D. Zeng, Y. Niu, Y. Gao, and D. Zhang, "Research on fault diagnosis of coal mill system based on the simulated typical fault samples," *Measurement*, vol. 161, Sep. 2020, Art. no. 107864, doi: 10.1016/j.measurement.2020.107864.

[6] P. F. Odgaard and B. Mataji, "Observer-based fault detection and moisture estimating in coal mills," *Control Eng. Pract.*, vol. 16, no. 8, pp. 909–921, Aug. 2008, doi: 10.1016/j.conengprac.2007.10.008.

[7] V. Agrawal, B. K. Panigrahi, and P. M. V. Subbarao, "A unified thermo-mechanical model for coal mill operation," *Control Eng. Pract.*, vol. 44, pp. 157–171, Nov. 2015, doi: 10.1016/j.conengprac.2015.08.002.

[8] V. Agrawal, B. K. Panigrahi, and P. M. V. Subbarao, "Intelligent decision support system for detection and root cause analysis of faults in coal mills," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 934–944, Aug. 2017, doi: 10.1109/TFUZZ.2016.2587325.

[9] I. Fagarasan and S. S. T. Iliescu, "Applications of fault detection methods to industrial processes," *WSEAS Trans. Syst.*, vol. 7, no. 6, pp. 812–821, 2008.

[10] E. Kisić, Ž. Durović, B. Kovačević, and V. Petrović, "Application ofT2Control charts and hidden Markov models in condition-based maintenance at thermoelectric power plants," *Shock Vibrat.*, vol. 2015, pp. 1–11, Jan. 2015, doi: 10.1155/2015/960349.

[11] Z.-G. Su, P.-H. Wang, X.-J. Yu, and Z.-Z. Lv, "Experimental investigation of vibration signal of an industrial tubular ball mill: Monitoring and diagnosing," *Minerals Eng.*, vol. 21, no. 10, pp. 699–710, Sep. 2008, doi: 10.1016/j.mineng.2008.01.009.

[12] Z. Qiao, X. Wang, H. Gu, Y. Tang, F. Si, C. E. Romero, and X. Yao, "An investigation on data mining and operating optimization for wet flue gas desulfurization systems," *Fuel*, vol. 258, Dec. 2019, Art. no. 116178, doi: 10.1016/j.fuel.2019.116178.

[13] H. Gu, H. Zhu, Y. Cui, F. Si, R. Xue, H. Xi, and J. Zhang, "Optimized scheme in coal-fired boiler combustion based on information entropy and modified K-prototypes algorithm," *Results Phys.*, vol. 9, pp. 1262–1274, Jun. 2018, doi: 10.1016/j.rinp.2018.04.045.

[14] Y. Lv, F. Fang, T. Yang, and C. E. Romero, "An early fault detection method for induced draft fans based on MSET with informative memory matrix selection," *ISA Trans.*, vol. 102, pp. 325–334, Jul. 2020, doi: 10.1016/j.isatra.2020.02.018.

[15] X. Han and X. Jiang, "Fault diagnosis of pulverizing system based on fuzzy decision-making fusion method 2 fault condition segmentation of pulverizing system based on grey," in *Fuzzy Information and Engineering*, vol. 2. Berlin, Germany: Springer, 2009, pp. 1–12.

[16] W. Qin, W. Yan, and J. Xu, "Application of fault diagnosis expert system in grinding process," in *Proc. IEEE Int. Conf. Autom. Logistics*, Aug. 2010, pp. 290–295, doi: 10.1109/ICAL.2010.5585295.

[17] J. Liu, L. Zhu, Y. Ma, J. Liu, W. Zhou, and D. Yu, "Anomaly detection of hot components in gas turbine based on frequent pattern extraction," *Sci. China Technol. Sci.*, vol. 61, no. 4, pp. 567–586, Apr. 2018, doi: 10.1007/s11431-017-9165-7.

[18] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometric Intell. Lab. Syst.*, vol. 171, pp. 16–25, Dec. 2017, doi: 10.1016/j.chemolab.2017.09.021.

[19] Y. Lv, X. Lv, F. Fang, T. Yang, and C. E. Romero, "Adaptive selective catalytic reduction model development using typical operating data in coal-fired power plants," *Energy*, vol. 192, Feb. 2020, Art. no. 116589, doi: 10.1016/j.energy.2019.116589.

[20] Q. Zhu, Q. Liu, and S. J. Qin, "Concurrent quality and process monitoring with canonical correlation analysis," *J. Process Control*, vol. 60, pp. 95–103, Dec. 2017, doi: 10.1016/j.jprocont.2017.06.017.

[21] S. Joe Qin, "Statistical process monitoring: Basics and beyond," *J. Chemometrics*, vol. 17, nos. 8–9, pp. 480–502, Aug. 2003, doi: 10.1002/cem.800.

[22] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593–1600, Jul. 2009, doi: 10.1016/j.automatica.2009.02.027.

[23] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring with kernel principal component analysis," in *Proc. Amer. Control Conf.*, Jun. 2010, pp. 7022–7027, doi: 10.1109/acc.2010.5531315.

[24] Z. Ge and Z. Song, "Distributed PCA model for plant-wide process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 5, pp. 1947–1957, Feb. 2013, doi: 10.1021/ie301945s.

[25] P. Tang, K. Peng, K. Zhang, Z. Chen, X. Yang, and L. Li, "A deep belief network-based fault detection method for nonlinear processes," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 9–14, 2018, doi: 10.1016/j.ifacol.2018.09.522.

[26] Y. Zhang, W. Du, Y. Fan, and L. Zhang, "Process fault detection using directional Kernel partial least squares," *Ind. Eng. Chem. Res.*, vol. 54, no. 9, pp. 2509–2518, Mar. 2015, doi: 10.1021/ie501502t.

[27] R. Eslamloueyan, "Designing a hierarchical neural network based on fuzzy clustering for fault diagnosis of the Tennessee–Eastman process," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 1407–1415, Jan. 2011, doi: 10.1016/j.asoc.2010.04.012.

[28] S. Ren, F. Si, J. Zhou, Z. Qiao, and Y. Cheng, "A new reconstruction-based auto-associative neural network for fault diagnosis in nonlinear systems," *Chemometric Intell. Lab. Syst.*, vol. 172, pp. 118–128, Jan. 2018, doi: 10.1016/j.chemolab.2017.12.005.

[29] R. Sharifi and R. Langari, "Nonlinear sensor fault diagnosis using mixture of probabilistic PCA models," *Mech. Syst. Signal Process.*, vol. 85, pp. 638–650, Feb. 2017, doi: 10.1016/j.ymssp.2016.08.028.

[30] V. N. Vapnik, "Statistical learning theory," *Encyclopedia Sci. Learn.*, vol. 41, no. 4, p. 3185, 1998.

[31] A. Kulkarni, V. K. Jayaraman, and B. D. Kulkarni, "Knowledge incorporated support vector machines to detect faults in Tennessee Eastman process," *Comput. Chem. Eng.*, vol. 29, no. 10, pp. 2128–2133, Sep. 2005, doi: 10.1016/j.compchemeng.2005.06.006.

[32] S. Mahadevan and S. L. Shah, "Fault detection and diagnosis in process data using one-class support vector machines," *J. Process Control*, vol. 19, no. 10, pp. 1627–1639, 2009, doi: 10.1016/j.jprocont.2009.07.011.

[33] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001, doi: 10.1162/15324430152748236.

[34] E. Zio and F. Di Maio, "Fatigue crack growth estimation by relevance vector machine," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10681–10692, 2012, doi: 10.1016/j.eswa.2012.02.199.

[35] A. Thayananthan, R. Navaratnam, B. Stenger, P. H. Torr, and R. Cipolla, "Pose estimation and tracking using multivariate regression," *Pattern Recognit. Lett.*, vol. 29, pp. 1302–1310, Jul. 2008, doi: 10.1016/j.patrec.2008.02.004.

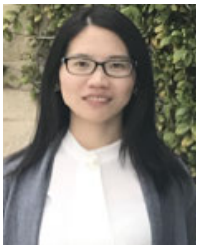
[36] Y. Ha and H. Zhang, "Fast multi-output relevance vector regression," *Econ. Model.*, vol. 81, pp. 217–230, Sep. 2019, doi: 10.1016/j.econmod.2019.04.007.

[37] C. Yuan and C. Neubauer, "Bayesian sensor estimation for machine condition monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. ICASSP*, vol. 2, Apr. 2007, pp. II-517–II-520, doi: 10.1109/ICASSP.2007.366286.

[38] B. Huang, "Bayesian methods for control loop monitoring and diagnosis," *J. Process Control*, vol. 18, no. 9, pp. 829–838, 2008, doi: 10.1016/j.jprocont.2008.06.006.
[39] M. E. Tipping, "Bayesian inference: An introduction to principles and practice in machine learning," in *Advanced Lectures on Machine Learning*, vol. 3176, 2004, pp. 41–62, doi: 10.1007/978-3-540-28650-9_3.
[40] J. Yu and S. J. Qin, "Multiway Gaussian mixture model based multiphase batch process monitoring," *Ind., Eng. Chem. Res.*, vol. 48, no. 18, pp. 8585–8594, 2009, doi: 10.1021/ie900479g.
[41] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models," *AIChE J.*, vol. 54, no. 7, pp. 1811–1829, 2008, doi: 10.1002/aic.
[42] J. Yu, "Engineering applications of artificial intelligence A new fault diagnosis method of multimode processes using Bayesian inference based Gaussian mixture contribution decomposition," *Eng. Appl. Artif. Intell.*, vol. 26, no. 1, pp. 456–466, 2013, doi: 10.1016/j.engappai.2012.09.003.
[43] Q. Jiang, B. Huang, and X. Yan, "GMM and optimal principal components-based Bayesian method for multimode fault diagnosis," *Comput. Chem. Eng.*, vol. 84, pp. 338–349, Jan. 2016, doi: 10.1016/j.compchemeng.2015.09.013.
[44] S. W. Choi, E. B. Martin, A. J. Morris, and I.-B. Lee, "Fault detection based on a maximum-likelihood principal component analysis (PCA) mixture," *Ind. Eng. Chem. Res.*, vol. 44, no. 7, pp. 2316–2327, Mar. 2005, doi: 10.1021/ie049081o.
[45] M. Quiñones-Grueiro, A. Prieto-Moreno, C. Verde, and O. Llanes-Santiago, "Data-driven monitoring of multimode continuous processes: A review," *Chemometric Intell. Lab. Syst.*, vol. 189, pp. 56–71, Jun. 2019, doi: 10.1016/j.chemolab.2019.03.012.
[46] X. Jiang, H. Zhao, and B. Jin, "Multimode process monitoring based on sparse principal component selection and Bayesian inference-based probability," *Math. Problems Eng.*, vol. 2015, pp. 1–12, Jan. 2015, doi: 10.1155/2015/465372.
[47] S. Tan, F. Wang, J. Peng, Y. Chang, and S. Wang, "Multimode process monitoring based on mode identification," *Ind. Eng. Chem. Res.*, vol. 51, no. 1, pp. 374–388, Jan. 2012, doi: 10.1021/ie102048f.
[48] J. Liu and D. Chen, "Nonstationary fault detection and diagnosis for multimode processes," *AIChE J.*, vol. 59, no. 4, pp. 215–228, 2012, doi: 10.1002/aic.11999.
[49] H. Kodamana, R. Raveendran, and B. Huang, "Mixtures of probabilistic PCA with common structure latent bases for process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 2, pp. 838–846, Mar. 2019, doi: 10.1109/TCST.2017.2778691.
[50] J. Pang, D. Liu, Y. Peng, and X. Peng, "Multiple-output-Gaussian-process regression-based anomaly detection for multivariate monitoring series," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 326–332, doi: 10.1109/PHM-Chongqing.2018.00062.
[51] M. A. T. Figueiredo, S. Member, and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 1–16, Aug. 2002.
[52] T. Andriyas and S. Andriyas, "Use of multivariate relevance vector machines in forecasting multiple geomagnetic indices," *J. Atmos. Solar-Terr. Phys.*, vol. 154, pp. 21–32, Feb. 2017, doi: 10.1016/j.jastp.2016.11.002.
[53] F. Shumin, L. Yanhong, and C. Lin, "Fault diagnosis based on RSPNN for pulverizing systems," *Control Eng. China*, vol. 19, no. 3, pp. 412–415, 2012, doi: 10.14107/j.cnki.kzgc.2012.03.016.
[54] A. Cortinovis, M. Mercangöz, T. Mathur, J. Poland, and M. Blaumann, "Nonlinear coal mill modeling and its application to model predictive control," *Control Eng. Pract.*, vol. 21, no. 3, pp. 308–320, Mar. 2013, doi: 10.1016/j.conengprac.2012.10.006.

WEI FAN received the B.S. degree in thermal and power engineering from the Nanjing Institute of Technology, Jiangsu, China, in 2015. He is currently pursuing the Ph.D. degree in power engineering and engineering thermophysics with Southeast University, Nanjing, Jiangsu. His research interests include process monitoring, fault detection, and diagnosis of complex industrial process, which brings together data-driven process modeling, soft sensors, multivariate analysis, and optimization.

SHAOJUN REN received the B.S. degree in thermal and power engineering and the Ph.D. degree in power engineering and engineering thermophysics from Southeast University, Nanjing, Jiangsu, China, in 2012 and 2018, respectively. His research interests include big data analysis of power systems, system characteristic modeling and fault diagnosis of large thermal power units, and so on.



QINQIN ZHU (Member, IEEE) received the Ph.D. degree from the Chemical Engineering Department, University of Southern California, advised by Prof. Joe Qin. She worked as a Senior Research Scientist with Facebook Inc., USA. Her research addresses theoretical challenges and problems of practical importance in the area of process systems engineering. By leveraging the power of mathematical modeling and optimization, her group strives to develop advanced multivariate statistical analysis algorithms that enhance decision making in complex engineering systems. Her research interests include developing advanced statistical machine learning methods, process data analytics techniques and optimization algorithms in the era of big data with applications to statistical process monitoring and fault diagnosis.

ZHIJUN JIA received the B.S. degree in thermal and powering engineering from the Inner Mongolia University of Technology, Inner Mongolia, China, in 2006, and the M.S. degree in power engineering and field engineering from North China Electric Power University, Beijing, China, in 2014.

DELONG BAI received the B.S. degree in production process automation from North China Electric Power University, Beijing, China, in 1997, and the M.S. degree in electrical engineering from Tsinghua University, Beijing, in 2014.

FENGQI SI received the Ph.D. degree in power machinery and engineering from Southeast University, in 2001. From October 2007 to October 2008, he was a Visiting Scholar with Lehigh University, USA. He is currently an Associate Dean of the School of Energy and Environment, Southeast University. His main research interests include the operating characteristics modeling of large thermal power units, thermal automatic control, optimization, and fault diagnosis of industrial processes.