Research Article

Multimodel Anomaly Identification and Control in Wet Limestone-Gypsum Flue Gas Desulphurization System

Xiaoli Li ^{1,2,3} Quanbo Liu ¹ Kang Wang ¹ Fuqiang Wang⁴ Guimei Cui,⁵
and Yang Li ⁶

¹Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

²Beijing Key Laboratory of Computational Intelligence and Intelligent System,
Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China

³Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China

⁴Technology Research Center, Shenhua Guohua(Beijing) Electric Power Research Institute Corporation, Beijing 100025, China

⁵School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China

⁶School of International Studies, Communication University of China, Beijing 100024, China

Correspondence should be addressed to Xiaoli Li; lixiaolibjut@bjut.edu.cn

Received 24 June 2020; Accepted 3 August 2020; Published 12 October 2020

Guest Editor: Jing Na

Copyright © 2020 Xiaoli Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sulphur dioxide, as one of the most common air pollutant gases, brings considerable numbers of hazards on human health and environment. For the purpose of reducing the detrimental effect it brings, it is of urgent necessity to control emissions of flue gas in power plants, since a substantial proportion of sulphur dioxide in the atmosphere stems from flue gas generated in the whole process of electricity generation. However, the complexity and nondeterminism of the environment increase the occurrences of anomalies in practical flue gas desulphurization system. Anomalies in industrial desulphurization system would induce severe consequences and pose challenges for high-performance control with classical control strategies. In this article, based on process data sampled from 1000 MW unit flue gas desulphurization system in a coal-fired power plant, a multimodel control strategy with multilayer parallel dynamic neural network (MPDNN) is utilized to address the control problem in the context of different anomalies. In addition, simulation results indicate the applicability and effectiveness of the proposed control method by comparing with different cases.

1. Introduction

With the increasing emergence of hazy weather over the past few years, environmental deterioration has raised great concern in China. Sulphur dioxide, serving as one of the primary contributors giving rise to atmospheric pollution, has a negative effect on human health as well as the environment [1–3]. Moreover, the combustion of fossil fuels, especially coal, in power plants is the major cause for existence of sulphur dioxide in the atmosphere [4]. Therefore, the removal of sulphur dioxide in flue gas from coal-fired power plants is of great significance and has gradually become the overriding concern to the public. Among a wide variety of flue gas desulphurization methods, wet limestone-

gypsum flue gas desulphurization (WFGD) is most power plants' first priority due to the fact that it is characterized by low operation cost and relatively high utilization of absorption reagent [5, 6]. In order to suppress the pollution brought by sulphur dioxide, the concentration of outlet flue gas along with a slurry pH value are selected as the controlled variables in our case study, and the former of which is a direct indicator that reflects desulphurization performance. Also, it has been verified that the pH value in slurry tank has a major influence on desulphurization efficiency [7].

Anomaly detection (or outlier detection) is the process of identifying unexpected observations or events, which differ significantly from the majority of the data [8]. Due to the fact that anomaly detection is an area of high application value, it

has been extended to various fields, especially in the industry. In this work, we focus on the case where flue gas desulphurizes in the context of anomalous conditions. It is well established that industrial processes and facilities therein are of ever-increasing complexity over the past few years, which results in a greater chance for occurrences of distinct faults. Moreover, given that flue gas desulphurization is a type of chemical engineering process, there exist some unique faults that induce anomalous conditions other than traditional industrial ones, such as catalyst deactivation, valve blockage, or compressor failure [9]. Thus, regarding chemical systems, a special attention must be placed on reliability and stability of systems during the whole operation, which entails detecting anomalous conditions and performing effective control on the system. In the past several decades, there are quite a few publications covering the subject of anomaly detection. However, there are few publications covering anomaly detection in the desulphurization system, let alone apply control techniques in anomalous conditions therein. To the best of our knowledge, nearly all of the contributions are concerned with fault detection and diagnosis problems. In literature [10, 11], principal component analysis (PCA) with monitoring indices Hotelling's T2 and squared prediction error is utilized to find the exact fault appearance during a specified time-frame in flue gas desulphurization system. The difference in kernel PCA is applied in the latter work, and the hybrid prediction model, which contains autoregressive integrated moving average model and least squares support vector machine model, is suggested in [12] to make prediction of the slurry pH value in the first place, and then a fault diagnosis system is designed to locate faults of the pH sensor. Additionally, modelling based on physical characteristics of combustion in flue gas desulphurization is introduced in [13]. In this method, a flue gas estimation model is built and used to identify process anomalies together with sensor malfunctions.

As for publications mentioned above, there exists a commonality that the realization of fault detection or diagnosis with proposed methods is driven by labeled instances, namely, all instances used contain explicit labels of various anomalies. Nevertheless, from a practical point of view, it is challenging to group anomalies into multiple categories (i.e., faults of different types) accurately in WFGD system, as the malfunction of certain part in system may trigger successive faults in other components. Accordingly, the classification results depend critically upon individual experience in practice. In this case, the usage of exploring and grouping method is more advantageous for the extraction of underlying anomalous conditions in practical applications. Therefore, in this article, the integration of isolation forest and adaptive mean shift clustering are employed to determine types of anomalous conditions existing in our system. Another issue that has to be taken into account is what kind of control strategy should we apply in this context. Here, multimodel control is utilized in our case study due to complicity and diversity of operating environments in the whole desulphurization process. More specifically, as a fair number of devices and plants are

employed on-site, there are many possibilities for occurrences of anomalies, which are actually one of the major reasons for variations in the environment. Moreover, since many subprocesses in flue gas desulphurization are interactive, the presence of unknown and discontinuous anomalous conditions in any subprocess, even caused by some unnoticeable devices, may bring about parameter changes in controlled actual system. Given that multimodel control is particularly suited to cope with a class of control problems, which caused by parameter uncertainties due to changes of operating environments, it is selected as the control strategy in our case study.

The main contribution of this work can be summarized as follows: (1) a novel exploring and grouping technique is proposed to provide a detailed analysis on anomalous environments in the industrial field. (2) MPDNN-based adaptive control is proposed and extended to multimodel for the first time. (3) Different combinations of control modes with respect to the reinitialized MPDNN adaptive model are compared and discussed.

The remainder of this paper is organized as follows: Section 2 describes the principle and process of wet flue gas desulphurization briefly. Then, the integrated anomaly detection technique based on isolation and clustering algorithms is introduced in Section 3. Section 4 discusses theoretical knowledge on identification and control process of multimodel system with MPDNN. In Section 5, simulation studies show the efficacy of the presented control strategy by comparing against numerical cases. Finally, Section 6 concludes this paper and our future work is given.

2. The Introduction of Flue Gas Desulphurization System

In fact, the whole process of sulphur dioxide abatement in WFGD system is complicated, although the chemical equation is seemingly simple from its form. The overall reaction of WFGD can be expressed in terms of a chemical equation as follows:

The basic reaction principle can be obtained from the overall reaction equation above. That is, the limestone in slurry serves as an absorbent reacting with sulphur dioxide in flue gas, and CO_2 along with gypsum ($\text{CaSO}_4 \cdot \text{H}_2\text{O}$) is finally produced. In the following, the process of wet flue gas desulphurization is outlined. For purposes of clarity, the general structure of WFGD system is presented in Figure 1, which provides an illustration for procedures in desulphurization. The limestone slurry as the absorption agent is fed to the absorber from slurry tank. In this process, both density and flux of the slurry can be measured via a densimeter and flowmeter mounted on the duct. At the same time, circulating pumps are utilized to rise up the slurry so that it can drop down from the top of the absorber. On the contrary, flue gas passes through the heater exchanger before entering the absorber from the inlet duct in order to lower

FIGURE 1: Schematic diagram of wet flue gas desulphurization system.

gas temperature. As the inlet duct is situated at the lower part of the absorber, it would result in countercurrent contact between the flue gas and limestone slurry. Next, consequential chemical reaction between two components removes sulphur dioxide content in flue gas. Finally, desulphurized flue gas discharges from the stack and gypsum, as a solid byproduct, is obtained in absorbent slurry, which can be applied in fields such as wallboard and cement [14, 15].

3. Extraction of Anomalous Conditions in WLFGD System

During the operation of WLFGD system, the operating data with an array of measured variables was collected from a 1000 MW coal-fired power plant unit every minute. In this section, some preliminary analyses of relevant variables are performed in the first place, which lays a foundation for anomalous conditions extraction later in this section. Then, relevant techniques utilized in the process of seeking anomalies are detailed.

3.1. Preliminaries. Here, four output variables in the system are selected to perform analysis on anomalies. They are outlet O_2 concentration, outlet sulphur dioxide concentration, outlet flue gas temperature, and slurry pH value. Before searching for anomalous conditions, the Pearson correlation coefficients between variables are calculated, and they are visualized in the form of correlation matrix (Figure 2).

The correlation matrix is featured by symmetry, unit diagonal, and semidefiniteness [16], each entry in the matrix stands for the correlation coefficient between two variables. The matrix mentioned above explicitly indicates the degree of relationship between each pair of variables. For notational simplicity, each variable mentioned above is denoted by parts of its full name in the obtained correlation matrix. As it is shown in Figure 2, the absolute values of Pearson correlation coefficients regarding four variables are all over 0.50, some of which even exceed 0.92, which reflects high levels of correlation among four output variables in WLFGD system. By [17], variables with high correlation imposes restrictions

FIGURE 2: Correlation matrix of four output variables.

on the performance of anomaly detection, for a minor deviance from normal scope of correlated variables may induce relatively greater anomaly. Moreover, high-dimensional data would bring about the problem of curse of dimensionality [18, 19]. PCA, as one of the most commonly used feature extraction techniques, reduces dimension by generating fewer principal components (PCs) to substitute for original variables. On the contrary, PCs are uncorrelated to each other [20]. In this sense, the problems arose by high-dimensionality data and correlated variables in anomalous conditions extraction can be addressed by means of PCA. Furthermore, the number of components to retain is another problem that needs to be taken into account before dimension reduction. The determination of PCs' number can refer to criteria as follows [21]: (1) eigenvalues of each PC should be greater than one; (2) each component explains at least 5% of variance; and (3) cumulative explained variance reaches 70%. Here, the summary of eigenvalues, proportion of explained variance, and cumulative explained variance for PCs with different numbers calculated by above four output variables are tabulated in Table 1.

As is shown in Table 1, the first two PCs have eigenvalues 3.068 and 1.247, yet the eigenvalue of the third PC is only 0.465. At the same time, the variance explained by the first

TABLE 1: Eigenvalues and variance for PCs.

Number	Eigenvalues	Explained variance proportion	Cumulative explained variance
1	3.068	0.623	0.623
2	1.247	0.253	0.876
3	0.465	0.094	0.970

two PCs is up to 62.3% and 25.3%, respectively, and cumulative explained variance reaches 87.6%. Based on the above-stated criteria, the retention of two PCs is preferable. Consequently, foregoing four output variables are reduced to two principle components for further analysis on anomalous conditions below.

3.2. Extraction of Anomalous Conditions. Having reduced dimensionality of practical measured data in WLFGD system, anomalous conditions would be detected and extracted thereafter, which can make necessary preparation for the construction of the multimodel control system in the next section. In this subsection, the integration of isolation-based and clustering-based method is employed to perform the task of anomalous conditions extraction in the context of WLFGD.

From a practical point of view, computation efficiency is of crucial importance to anomalies detection, especially in industrial scenarios, due to the requirements of initiating prompt action to maintain system stability. Isolation forest (IF) is a type of ensemble-based anomaly detection method proposed by Liu et al. in [22], which is characterized by high accuracy, low memory cost, and computation complexity [23], thus can guarantee the timeliness of searching for anomalies in system. Furthermore, for isolation forest algorithm, anomalous instances are not required during the training process [24]. Therefore, it is well-suited to implement in an industrial environment. The rationale of using IF to detect anomalies is on the basis of inherent properties of anomalies, that is, the quantity of deviated data is small and they are in stark contrast to normal ones. On these bases, provided that a tree structure is utilized to describe the recursive process of data space partitioning, anomalous data are more likely to be separated thus in the upper position of the tree compared with normal points. More specifically, let $\{x_n\}_{n=1}^N \subset \mathcal{R}^d$ be the points of interest, where N and d denote the number of instances and attributes, respectively; the procedure for formation of an isolation forest is summarized as follows:

- (1) A subset is built by selecting k ($k < N$) instances randomly from N instances, and an *iTree* would be constructed by partitioning this subset recursively.
- (2) Firstly, select an attribute from the subset arbitrarily. Then, a split point is chosen as any point within the selected attribute in this subset.
- (3) The data space constituted by the subset is segmented into two subspaces, which correspond to left and right subtrees, by the chosen split point. The data would then be assigned to corresponding subtrees followed by comparing their values with the split

point in the specified attribute, and the smaller and greater ones correspond to left and right subtrees, respectively.

- (4) Repeat the second and third steps for each new subspace until there is only a single point in the subspace, or the height limit h of *iTree* is fulfilled. h is a parameter depending on the size of the subset with the form of $h = \lceil \log_2^k \rceil$. In this way, an *iTree* is established.
- (5) Constructing n *iTree* according to steps stated above such that the isolation forest $\{T_1, \dots, T_n\}$ can be formed.

As for any point x , the partition time to isolate that point in i th *iTree* can be referred to as path length $L(x_i)$, which provides reference for judging whether x is an anomaly. In this case, n path lengths can be acquired, and each of which corresponds to an *iTree* in $\{T_1, \dots, T_n\}$. Then, the path length for x in an isolation forest is calculated by taking the average of n path lengths, i.e.,

$$E(L(x)) = \frac{1}{n} \sum_{i=1}^n L(x_i). \quad (2)$$

The anomaly score that measures the outlieriness of x is defined in [17] and given as

$$s(x) = 2^{((E(L(x)))/(c(k)))}, \quad (3)$$

where k is the number of instances in the subset, and $c(k)$ is used to normalize $L(x)$ and can be written in the form

$$c(k) = 2 * \left[H(k-1) - \frac{(k-1)}{k} \right], \quad (4)$$

with $H(x)$ denoting the harmonic number, which can be estimated by Euler's constant as $H(x) = \ln(x) + 0.5772156649$. Taking two extreme cases into account, where the minimum and maximum distances from x to the root of each full-grown *iTree* are achieved, respectively. At this point, $E(L(x))$ reaches the minima 0 and maxima $n-1$. Substituting two extreme values of $E(L(x))$ into (3), it can be concluded that the anomaly score for x is within the range $[0, 1]$. Once the anomaly score has been calculated for each sample, anomalies can be sought out based on the threshold of anomaly score, and the more it approximates to 1, x is more likely to be an anomaly. As a rule of thumb, the threshold of anomaly score is taken as 0.6. In our case study, the instances contained in the subset and the number of trees are set as 256 and 100, respectively, which have been proven to be the optimal parameters that

yield the best performance of anomaly detection [22]. On this basis, anomalous points were searched among dimension-reducing operating data. To simplify the presentation, the result obtained in a certain period of time is given in Figure 3.

As shown in Figure 3, partial detected anomalies are marked by red points and remaining points are considered as normal ones. Likewise, anomalous points over different periods of time can be collected and pooled together. Although all anomalies have been found at this stage, there is a need to determine the type of anomalous condition for further study. However, due to the complexity of operating conditions in practical WLFGD scenarios, it is challenging to know the specific number of anomalous conditions, and the rational partition of them is also an intractable problem. Here, adaptive mean shift is applied to address the encountered problem.

Mean shift (MS) is a nonparametric mode seeking algorithm based upon kernel density estimate, which is able to find clusters with any shapes in data [25]. The major downside to MS is its high computational complexity in high dimensions [26]. Nevertheless, in our case study, the dimension of feature space has been reduced in advance, which avoids the computational problem it brings. Meanwhile, the number of modes (i.e., clustering numbers) can be automatically determined by the algorithm; therefore, MS is of practical significance for determining the type of anomaly in our case study. The principle of MS clustering is outlined as the process of moving each point towards the direction of maximum local density, and the direction is determined via kernel density estimate. The movements of each point are repeated until the point converges to another point with local maxima of density (i.e., mode). Lastly, points with the same mode are classified as the identical cluster. The bandwidth of a kernel function is the unique and key parameter in MS which affects the clustering result of the MS significantly [27, 28]. Based on whether the bandwidth is variable during the shifting of points, MS can be divided into two groups, i.e., constant and adaptive mean shifts. By for constant bandwidth, either too small or too large bandwidths may lead to inappropriate clustering. By comparison, adaptive mean shift (AMS) can yield better results for clustering. Assume $x_i \in \mathbb{R}^d$, $i = 1, \dots, N$ be the data points in a d -dimensional feature space, and each of which is assigned with a bandwidth h_i . The density estimation of the sample point with kernel profile k is defined as

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N h_i^{-d} k\left(\left\|\frac{x - x_i}{h_i}\right\|^2\right), \quad (5)$$

where Gaussian kernel is used; the corresponding profile is a function of the form

$$k(\|x\|^2) = \exp\left(-\frac{1}{2}\|x\|^2\right). \quad (6)$$

FIGURE 3: Anomalies detected by IF in WLFGD system.

Let function $g(x) = -k'(x)$, then the gradient of kernel density estimation shown in (5) is written as

$$\nabla \hat{f}(x) = \frac{2}{N} \sum_{i=1}^N h_i^{-(d+2)} g\left(\left\|\frac{x - x_i}{h_i}\right\|^2\right) \frac{\left[\frac{\sum_{i=1}^N x_i h_i^{-(d+2)} g\left(\left\|((x - x_i)/h_i)\right\|^2\right)}{\sum_{i=1}^N h_i^{-(d+2)} g\left(\left\|((x - x_i)/h_i)\right\|^2\right)} - x\right]}{\quad} \quad (7)$$

where the rightmost part in (7) is known as the mean shift vector, whose direction yields maximum density increment. Each point in the feature space moves to local density maxima progressively with the direction of gradient, namely, the direction of mean shift vector. Now, there arises the question that how to determine the adaptive bandwidth h_i . Here, the k -nearest neighbor method in [26] is adopted. That is, the bandwidth of x_i is dependent on its k th distant point $x_{i,k}$, and h_i takes the form

$$h_i = \|x_i - x_{i,k}\|. \quad (8)$$

It follows from (8) that the value of k greatly influences the clustering outcome by changing the bandwidth of each point. On the basis of detected anomalous points, clustering is performed by adaptive mean shift with different k . Meanwhile, silhouette coefficient is introduced as the measure of clustering performance to select the optimal k value. As for the clustering silhouette coefficient, the value is in the range $[-1, 1]$ and the rationality of the clustering result is proportion to its value. With anomalies collected by IF, the silhouette coefficient is calculated and plotted against k with the interval of 50; the result is presented in Figure 4.

As is depicted in Figure 4, it is clear that the clustering silhouette coefficient reaches its maxima when k equals to 100. In this sense, the optimal clustering effect can be achieved with $k=100$. Accordingly, the subdivision of anomalous conditions in WLFGD system was carried out by AMS in this context. To enhance the accuracy of clustering, both dimensions of obtained anomalies are scaled to $[-1, 2]$ before clustering, and the result is shown in Figure 5.

As illustrated in Figure 5, anomalies are classified into three categories, each of which is constituted by dots in a certain color. Meanwhile, the mode of each category is

FIGURE 4: Clustering silhouette coefficient with different k .

FIGURE 5: Extraction result of different anomalous conditions.

marked by a solid circle, respectively. At this stage, the data of various anomalous conditions can be extracted via the clustering result mentioned above. Then, the dataset corresponding to each type of anomalous condition was randomly classified into two portions, including 70 percent for model training and 30 percent for control testing, and both of the training and control process would be made clear in the next section.

4. Identification and Multimodel Adaptive Control in Anomalous Conditions

In effect, anomalous conditions extraction done before makes provision for adaptive multimodel control introduced in this section. The overall control process can be briefly summarized as two steps. First, models are established to characterize different types of anomalous conditions extracted above. Then, the multimodel control strategy is applied in connection with those models. Above-mentioned two steps correspond to the following two subsections separately. Besides, as pointed out in the first part of this work, the pH values in limestone slurry and outlet sulphur dioxide concentration are chosen as

controlled variables (i.e., output variables) of the system. Second, the flux and density of limestone slurry, which are readily available through sensors meanwhile have a considerable impact on system outputs, are selected as control variables (i.e., input variables) of the system.

4.1. System Identification. Dynamic neural network has drawn increasing attention and been exploited in many fields, due to its preferable capabilities in representation of nonlinear system compared with the static one. Moreover, dynamic neural network used for nonlinear system identification can be further categorized as series-parallel model and parallel model [29]. In view of the fact that reactions taking place in the absorber have high level of complexity, thus there exist great obstacles to model the system with chemical mechanism. Meanwhile, high nonlinearity and delay existing in the process of slurry pH variations further add the difficulty for system identification in the context of WLFGD. On account of the abovementioned merits of dynamic neural network, here, the parallel form in it is utilized to identify the system in our case study. The strengths of this type of neural network can be summarized as follows: (1) the output of actual system is less affected by the bias resulting from noise [30], and (2) its identification model is particularly suitable for using offline. As the parallel dynamic neural network with no hidden layer has been described in detail in [31–33], in what follows, the description of MPDNN is our main concern.

Consider the nonlinear system to be identified is given as

$$\dot{x}_t = f(x_t, u_t, t), \quad (9)$$

where the state $x_t \in \mathfrak{R}^n$ and input $u_t \in \mathfrak{R}^m$, and the dynamic neural network of the following form is used to identify the system:

$$\hat{\dot{x}}_t = A\hat{x}_t + W_{1,t}\sigma(V_{1,t}\hat{x}_t) + W_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t), \quad (10)$$

with $\hat{x}_t \in \mathfrak{R}^n$ being the estimate of state. $A \in \mathfrak{R}^{n \times n}$ is a diagonal Hurwitz matrix. The control input function is assumed bounded and taken as $\gamma(u_t) = u_t$. Let l denote the number of nodes in the hidden layer, then $V_{1,t}$ and $V_{2,t}$ are l -by- n weight matrices and $W_{1,t}$ and $W_{2,t}$ are n -by- l weight matrices. $\sigma(\cdot) \in \mathfrak{R}^l$ is a sigmoidal vector function, and $\phi(\cdot) \in \mathfrak{R}^{l \times l}$ is a diagonal matrix constituted by elements of the sigmoidal vector function, which can be represented as $\phi(\cdot) = \text{diag}[\phi_1(V_{2,t}\hat{x}_t), \dots, \phi_l(V_{2,t}\hat{x}_t)]$, and the component of sigmoidal vector function takes the form

$$\phi_i(x) = \sigma_i(x) = \frac{a}{1 + \exp(-bx)} - c, \quad i = 1, 2. \quad (11)$$

In reality, nonlinear system (9) cannot be represented by (10) precisely due to the presence of unmodeled dynamics Δf , which can be formulated as

$$\Delta f = \dot{x}_t - Ax_t + W_1^* \sigma(V_1^* x_t) + W_2^* \phi(V_2^* x_t) \gamma(u_t), \quad (12)$$

with W_1^* , W_2^* , V_1^* , and V_2^* being any fixed weight matrices. Since the matrix A in (10) is stable, the pair $(A, R^{1/2})$ is controllable, the pair $(Q^{1/2}, A)$ is observable, and the local frequency condition expressed by (13) holds [34]

$$A^T R^{-1} A - Q \geq \frac{1}{4} [A^T R^{-1} - R^{-1} A] R [A^T R^{-1} - R^{-1} A]^T. \quad (13)$$

Then, the matrix Riccati equation of the form,

$$A^T P + P A + P R P + Q = 0, \quad (14)$$

has a positive symmetric solution P . Then, it follows from [35] that weight matrices update law can be given as follows:

$$\dot{W}_{1,t} = K_1 P \Delta_t \left(\sigma^T + \hat{x}_t^T \tilde{V}_{1,t}^T \frac{\partial \sigma^T(Z)}{\partial(Z)} \Big|_{Z=V_{1,t} \hat{x}_t} \right), \quad (15a)$$

$$\dot{W}_{2,t} = K_2 P \Delta_t \left(u_t^T \phi^T(V_{2,t} \hat{x}_t) + \hat{x}_t^T \tilde{V}_{2,t}^T \frac{\partial(\phi_t u_t)^T(Z)}{\partial(Z)} \Big|_{Z=V_{2,t} \hat{x}_t} \right), \quad (15b)$$

$$\dot{V}_{1,t} = K_3 P W_{1,t} \left(\frac{\partial \sigma^T(Z)}{\partial(Z)} \Big|_{Z=V_{1,t} \hat{x}_t} \right) \Delta_t \hat{x}_t^T - K_3 \Lambda_1 \tilde{V}_{1,t} \hat{x}_t \hat{x}_t^T, \quad (15c)$$

$$\dot{V}_{2,t} = K_4 P W_{2,t} \left(\frac{\partial(\phi_t u_t)^T(Z)}{\partial(Z)} \Big|_{Z=V_{2,t} \hat{x}_t} \right) \Delta_t \hat{x}_t^T - K_4 \Lambda_2 \tilde{V}_{2,t} \hat{x}_t \hat{x}_t^T, \quad (15d)$$

where the identification error Δ_t is denoted as $\hat{x}_t - x_t$. Gain matrices $\{K_i\}_{i=1}^4 \in \mathfrak{R}^{n \times n}$ are utilized to determine the learning rate of neural network. Both Λ_1 and Λ_2 are positive definite matrices, and P stands for the solution of Riccati equation in (14). In addition, $\tilde{V}_{i,t} \triangleq V_i^* - V_{i,t}$ ($i = 1, 2$), and when $t = 0$, $V_{i,0} = V_i^*$. In our case study, flux F_t and density D_t of limestone slurry are regarded as inputs of the neural network. On the contrary, outlet sulphur dioxide concentration $x_{1,t}$ and slurry pH value $x_{2,t}$ are quantities to be estimated. Thus, the input-output mapping relation can be given as

$$[x_{1,t}, x_{2,t}]^T = f(F_t, D_t), \quad (16)$$

where f is a nonlinear vector-valued function. Aimed at enhancing the convergence speed, all measurements are scaled to zero mean and unit variance before the online identification. In the following, system identification is carried out with or without the existence of hidden layer in neural network. Further, in the case where the hidden layer is involved, two and three nodes are used, respectively, in order to find the form with the optimal identification accuracy. Here, due to the space limitation, only one type of extracted anomalous condition is chosen to be identified. As for the neural network without hidden layer, $\phi(\cdot)$ and weight matrices $V_{1,t}$, $V_{2,t}$ in (10) are set as the

identity matrix. Sigmoidal functions in all three cases are given as

$$\phi_i(x) = \sigma_i(x) = \frac{1.2}{1 + \exp(-0.5x)} - 0.1, \quad i = 1, 2. \quad (17)$$

Initial weight matrices are given as

$$W_{1,0} = \begin{bmatrix} 0.25 & 1 \\ 1 & 0.25 \end{bmatrix}, \quad (18)$$

$$W_{2,0} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.2 \end{bmatrix}.$$

Additionally, $A = \begin{bmatrix} -20 & 0 \\ 0 & -10 \end{bmatrix}$, $R = \begin{bmatrix} 15 & 0 \\ 0 & 10 \end{bmatrix}$, and $Q = \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix}$. Thus, the solution P of (14) can easily be calculated as $P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. Adaptive gains are selected as

$K_1 = K_2 = 15$ and the initial states are set as $\hat{x}_0 = [0, 0]^T$. It should be noted that the selection of A , R , and Q should preferably make sure that the solution P is a real matrix such that computation time in the identification process can be reduced. Moreover, adaptive gains can be appropriately increased to speed up the identification process. On this base, identification for anomalous condition in the selected training set is obtained by this no-hidden-layer dynamic neural network with weight updating law given in [32]. The estimates and actual outputs are given simultaneously in Figure 6 for comparison purpose.

As for MPDNN with one hidden layer, output weights are the same as $W_{1,0}$ and $W_{2,0}$ set above, and input weight matrices in the hidden layer are initialized as

$$V_{1,0} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}, \quad (19)$$

$$V_{2,0} = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}.$$

In such a case, the system is identified by MPDNN in (10) with two nodes in the hidden layer, weights are updated based on (15) with $K_i = 15$ ($i = 1, 2, 3, 4$), $\Lambda_1 = \Lambda_2 = I$, while the setting of remaining parameters remain unchanged. Figure 7 presents the identification result for both states.

Likewise, in the case where three nodes are in the hidden layer, initial weight matrices are given as follows:

$$W_{1,0} = W_{2,0} = \begin{bmatrix} 0.1 & 0 & 0.2 \\ 0 & 0.2 & 0.5 \end{bmatrix}, \quad (20)$$

$$V_{1,0} = V_{2,0} = \begin{bmatrix} 0 & 0.3 & 0.2 \\ 0.3 & 0.5 & 0 \end{bmatrix}^T.$$

Remaining parameters are selected the same as those in the former two cases; the identification results are shown in Figure 8. It should be mentioned that the excessive number

FIGURE 6: Identification results of dynamic neural network with no hidden layer: (a) result for outlet sulphur dioxide concentration; (b) result for the pH value in slurry.

FIGURE 7: Identification results of dynamic neural network with two nodes in the hidden layer: (a) result for outlet sulphur dioxide concentration; (b) result for the pH value in slurry.

of parameters in weight matrices may pose limitations on identification speed. Therefore, certain elements in the abovementioned weight matrices are set as zero and unchanged during the identification process.

Next, identification is conducted for the testing set with the same parameters. Due to the space limitation, results are not visualized here. Instead, the identification performance of three kinds of neural networks is measured by the mean squared error (MSE) criterion, and corresponding results in training and testing sets are presented in Table 2.

From Table 2, it can be easily concluded that MPDNN outperforms the no-hidden-layer one both in the training set and testing set. More specifically, the one with three nodes in the hidden layer possesses the optimal identification performance. Due to the fact that the control scheme designed in the following section is based on the principle of certainty equivalence, thus the control performance depends critically on identification accuracy of models. On this base, MPDNN with three nodes in the hidden layer is selected to model other two types of anomalous conditions based upon their

FIGURE 8: Identification results of dynamic neural network with three nodes in the hidden layer: (a) result for outlet sulphur dioxide concentration; (b) result for the pH value in slurry.

TABLE 2: Identification results in training and testing sets.

Training set			Testing set		
Type	MSE for x_1	MSE for x_2	Type	MSE for x_1	MSE for x_2
No hidden layer	$1.898e-04$	$1.185e-05$	No hidden layer	0.0375	$3.457e-05$
Two nodes	$2.616e-05$	$1.435e-06$	Two nodes	$9.553e-03$	$1.002e-05$
Three nodes	$9.618e-06$	$7.993e-08$	Three nodes	$1.042e-03$	$8.921e-06$

respective training sets, and they would serve the control purpose below.

4.2. Control Design. As discussed in the introductory section, it is truly difficult to exert effective control over the whole desulphurization process; this can be attributed to the fact that actual operating environments in WLFSD system are complex and manifold. All the above considerations suggest that the adaptive multimodel control strategy is particularly suited to cope with such kind of problem, where the nonlinear system with large parametric uncertainty requires to be controlled under a time-varying environment. However, there also exist some issues precluding the application of multimodel control in practical scenarios, and we shall discuss them in the following.

To simplify the exposition, some mathematical notations are introduced into the statement of the problem. Let p be the parameter vector of the plant at a certain instant, while $\hat{p}_i$ denotes the parameter of the i th model among multiple models at the same instant. It is assumed that both p and $\hat{p}_i$ lie within a d -dimensional compact set U in the parameter space. Our objective is to estimate by means of multiple models at each instant. As for the i th fixed model, parameter vector $\hat{p}_i$ is located in the same position at all times. Rather, an adaptive model would approximate p progressively by parameter adjustment from initial values, hence parameters

of adaptive models are time-varying before convergence is reached. By far, a great deal of research has been conducted, both in linear and nonlinear cases, to find the optimal implementation of multimodel control, and distinct combinations of models are studied, which include (1) fixed models only; (2) adaptive models only; (3) fixed models and free running adaptive models; and (4) fixed models, reinitialized and free running adaptive models. It has been demonstrated that the last control architecture, which is of interest in this work, could yield the best performance, and the reader can refer to [36, 37] for more details. With this architecture, as long as one of the fixed models is selected, parameters of the reinitialized adaptive model would be reset to those of the selected models, and then adaptation of parameters would proceed from initialized ones. Provided that p is located somewhere in U at a certain instant, the neighborhood of $\{\hat{p}_i\}_{i=1}^N$ can be viewed as a class $\{U_i\}_{i=1}^N$ distributing in U . In this case, fixed models are served as providing initial parameter estimates for p such that it is included by a certain U_i . Then, tuning is carried out based on $\hat{p}_i$, namely, a point in U_i , to asymptotically converge to the real parameter p . The two stages in this process can eliminate the transient and steady-state error, respectively. Obviously, whether the control architecture of this sort can be applied effectively largely depends on the closeness between parameter $\hat{p}$ of fixed models (i.e., initial parameters of the adaptive model at switching instant) and p of the real plant at

a certain instant. In fact, large initial parametric error of the adaptive model would result in poor transient response, and long time may be required for adaptation and settling, which restricts the use of multimodel adaptive control in practice to a large extent. As a consequence, appropriate choices of $\hat{p}$ for different fixed models, which are the prerequisite for effective tuning in the subsequent stage, are of utmost significance.

Apart from the qualitative problem for determination of initial parameters, a quantitative problem concerning the number of fixed models is also formidable. The latter problem imposes limitations on practical application of such kind of control strategy as well, for computational overhead and efficiency that deal with the number of models must be taken into account. From a practical point of view, it is often the case that plants only switch among several subsets in U . Thus, there is no point in building fixed models with improper parameters. To handle two above-stated problems, sufficient knowledge about U should be gained such that parameters of fixed models are obtained; further, each subset U_i (i.e., the corresponding area in which adaptation of parameters takes place) could be determined. To this end, as indicated earlier, anomalous conditions were analyzed and extracted in turn and then subdivided into varied types. In this sense, information regarding anomalous conditions on-site can be obtained thoroughly, based on which the number of fixed models is therefore determined. Our idea is illustrated intuitively in the form of a flow diagram (Figure 9).

It can be seen clearly from Figure 9 that the multimodel control process is partitioned into two parts in our work, where the left-half part makes detailed analysis for different types of anomalous conditions by data exploring and mining. In this way, qualitative and quantitative information on anomalies is acquired, which could be further utilized to provide guidance for model set building in the right-half part of Figure 9. At the stage of control design, a corresponding controller is designed for each constructed model and the controller set $\{C_i\}_{i=1}^N$ is established. In fact, models in the model set $\{I_i\}_{i=1}^N$ and controllers in $\{C_i\}_{i=1}^N$ are pairwise. That is, one pair in $\{I_i\}_{i=1}^N$ is selected at each instant based on the switching scheme introduced later.

As stated in the introductory section, both controlled variables, slurry pH value, and outlet flue gas concentration are significant indicators that can measure the effectiveness of desulphurization. In our case study, a regulation problem would be resolved, i.e., multimodel adaptive control was carried out to track the reference values we set. The set point of slurry pH value should make trade-off between desulphurization efficiency and safety problems caused by corrosiveness. It has been verified in [38] that the reasonable value of pH is 5.5 at normal operation conditions. Besides, the desired outlet flue gas concentration is set as 3.6 mg/m³ in light of the practical condition and emission standard. Hereafter, the desired states are represented as $x^* = [x_1^*, x_2^*]^T$ and the control error vector is written as $e = x^* - \hat{x}$. Differentiating e with respect to t such that dynamics of the control error can be derived, which is expressed as

$$\dot{e}_t = \dot{x}_t^* - \dot{\hat{x}}_t, \quad (21)$$

and then (10) is invoked to rewrite (21) in the following form:

$$\dot{e}_t = \dot{x}_t^* - \dot{\hat{x}}_t = -A\hat{x}_t - W_{1,t}\sigma(V_{1,t}\hat{x}_t) - W_{2,t}\phi(V_{2,t}\hat{x}_t)\gamma(u_t). \quad (22)$$

If we choose the control input to be in the form

$$\gamma(u_t) = u_t = -[W_{2,t}\phi(V_{2,t}\hat{x}_t)]^+ [Ax_t^* + W_{1,t}\sigma(V_{1,t}\hat{x}_t)], \quad (23)$$

where the first term on the right-hand side of (23) denotes Moore–Penrose inverse of matrix $W_{2,t}\phi(V_{2,t}\hat{x}_t)$.

Assumption 1. The number of nodes l in the hidden layer and the number of states n satisfy $l \geq n$. Furthermore, n -by- l matrix $W_{2,t}\phi(V_{2,t}\hat{x}_t)$ has full row rank, i.e., $r[W_{2,t}\phi(V_{2,t}\hat{x}_t)] = n$, such that the product of $W_{2,t}\phi(V_{2,t}\hat{x}_t)$ and its Moore–Penrose inverse equals to the identity matrix. In particular, $W_{2,t}\phi(V_{2,t}\hat{x}_t)$ is a square matrix when $l = n$, then the condition becomes nonsingularity of the matrix $W_{2,t}\phi(V_{2,t}\hat{x}_t)$ [39, 40].

By assumption, substituting (23) into (22), the closed-loop dynamics of error is obtained:

$$\begin{aligned} \dot{e}_t &= -A\hat{x}_t - W_{1,t}\sigma(V_{1,t}\hat{x}_t) \\ &\quad + W_{2,t}\phi(V_{2,t}\hat{x}_t)[W_{2,t}\phi(V_{2,t}\hat{x}_t)]^+ [Ax_t^* + W_{1,t}\sigma(V_{1,t}\hat{x}_t)] \\ &= A(x_t^* - \hat{x}_t) = Ae_t. \end{aligned} \quad (24)$$

Clearly, as A is a Hurwitz matrix, the control error is asymptotically stable with the control input we implemented. In addition, the switching time and sequence of multiple models are entirely dependent on the switching scheme. The switching scheme is designed on the basis of the performance index for each model. For the assurance of switching accuracy, both instantaneous and cumulative identification errors are required to be taken into account in performance index $\{J_i(t)\}_{i=1}^N$. The performance index chosen in our study has the form:

$$J_i(t) = \alpha e_i^2(t) + \beta \int_0^t e^{-\lambda(t-\tau)} e_i^2(\tau) d\tau, \quad (25)$$

with α and β being the weights of instantaneous and cumulative identification errors. $e_i(t)$ denotes the identification error at instant t , and the exponential weighting method is applied in the cumulative error calculation, wherein λ indicates the forgetting factor. The rationale for switching scheme is to search for $\min_i \{J_i(t)\}$ at each instance. The smaller performance index J_i is the higher identification accuracy of the i th model. On this base, the corresponding controller $C_i(t)$ is chosen such that the best control performance can be yielded at the corresponding instant.

FIGURE 9: Scheme of the complete control process.

5. Simulation Studies

In this section, different multimodel control modes are designed and analyzed on the basis of anomalous conditions extracted in previous sections. Moreover, MPDNN with three nodes in the hidden layer (without further specification, MPDNN denotes this kind of structure in this section) is used for model and controller design in the following cases.

5.1. Preparations. In previous sections, anomalous conditions in different types have been extracted, and one of which has been modelled via MPDNN over its training set. The convergent weights of neural network are recorded as θ_1 in this case. Similarly, the other two types of conditions can be identified over the respective training sets, during which the same technique (including the preprocessing of data) is used

and requires suitable choice of parameters in MPDNN. In this way, with the expected identification accuracy (MSE for both states in two cases is less than $2e-4$), three sets of parameter vectors were obtained.

$$\theta_j = (a, b, c, d, e, f, g, h), \quad j = 1, 2, 3, \quad (26)$$

and each a to h in θ are constituted by the corresponding pair of elements in W_i and V_i ($i = 1, 2$) as follows:

$$W_i = \begin{bmatrix} a_i & 0 & b_i \\ 0 & c_i & d_i \end{bmatrix}, \quad (27)$$

$$V_i = \begin{bmatrix} 0 & e_i & f_i \\ g_i & h_i & 0 \end{bmatrix}^T,$$

where

$$\begin{aligned} \theta_1 &= [(0.209, 0.259), (0.151, -0.298), (-0.703, 0.027), (0.713, -0.196), (-0.848, 0.52), (-0.52, -0.166), (-0.753, -0.9), (-0.632, 0.806)], \\ \theta_2 &= [(0.136, 0.533), (-0.848, 0.868), (-0.89, -0.748), (0.063, 0.138), (-0.061, 0.589), (-0.976, -0.378), (-0.326, -0.057), (-0.676, -0.669)], \\ \theta_3 &= [(-0.621, -0.835), (-0.252, 0.859), (0.262, 0.549), (0.571, -0.026), (-0.128, 0.022), (-0.106, 0.636), (-0.387, 0.59), (0.017, 0.289)]. \end{aligned} \quad (28)$$

The resulting weight matrices mentioned above are used to establish fixed models, respectively. To test and verify the effectiveness of the proposed method in practical WFLGD system, a simulation platform of anomalous conditions in our case study is set up. In this platform, each type of anomalous environment is simulated via MPDNN in the above form over respective testing data. With no loss of generality, normal condition is taken into account by fitting MPDNN to the part of the remaining normal operation data. To assure the accuracy of simulation, parameters for MPDNN in each situation were determined by trial and error such that the optimal identification performance could be achieved. In this way, the final mean squared identification errors for both states in all the four conditions are lower than $1.042e-3$. At this stage, both fixed models and simulation environments in WFLGD system have been established.

The platform constructed above would simulate both three types of anomalous conditions and the normal operation condition every 30 units of time, which can be expressed as

$$\begin{cases} 1 \text{ st Anomaly,} & 0 \leq t < 30, \\ 2 \text{ nd Anomaly,} & 30 \leq t < 60, \\ \text{Normality,} & 60 \leq t < 90, \\ 3 \text{ rd Anomaly,} & t \geq 90. \end{cases} \quad (29)$$

Remark 1. Initial weights of models have a major effect on the control performance of multimodel adaptive control, particularly in practical applications. The problem of interest in this paper is mainly concerned with the setting of initial weights in anomalous conditions; the discussion and analysis of normal operation conditions are out of scope. Hence, we do not arrange the testing set and build only one model for the normal condition.

The initial states of all MPDNN in the simulation platform are set as $x_0 = [0, 0]^T$. In addition, as for different cases conducted below, proper combinations of α, β , and λ were selected for the performance index in (21). In the following, simulations on different cases would be conducted in a sequence, with the objective of testing the effectiveness of MPDNN multimodel adaptive control with different combinations of fixed models.

5.2. Experiments. As mentioned previously, it has been verified that the mode of fixed models with reinitialized adaptive models has the optimal control effect in multimodel control. Moreover, the topic discussed in this work focuses more on the selection of fixed models in practical multimodel control. Therefore, combinations of different kinds of models would not be involved in cases below. Instead, varied methods for construction of fixed models are simulated to test the validity of the method proposed in this article. To simplify notations, in cases below, the i th MPDNN fixed model is denoted as IF_i , and the conventional and reinitialized MPDNN adaptive models are denoted as IA_1 and IA_2 , respectively.

Case 1. Single free running adaptive model:

The single identification model is the most basic form in adaptive control. The parameters of the model would adjust with variations of the environment. Adaptive control of this sort can be effective when slow variations take place in the environment, whereas the large overshoot would arise at the instant when unanticipated environment is met. The initial value of IA_1 in this case is set as θ_1 . Figure 10(a) displays the response of the plant in different conditions, and Figure 10(b) is the absolute value of the control error. Clearly, there exists large transient error when condition shifts at 30, 60, and 90 units of time. Worse still, after 90 s, strong oscillations arise and the plant is out of control, which reveals that the single adaptive model is not a suitable choice in practical WFLGD scenarios.

Case 2. Three time-based fixed models, one adaptive model, and one reinitialized adaptive model:

In this case, the building of $\{IF_i\}_{i=1}^3$ is dependent on the partition of time series. To be specific, all extracted anomalous points were ordered in ascending time sequence. Then, points were equally divided into three portions, and each of which are utilized to set up a fixed model in the same fashion as before with convergence weights and proper parameters, and mean squared identification errors for two states in all the three models are below $8.703e-4$. As for IA_1 applied in this case, which aims at controlling the plant in normal conditions, its initial parameters are set close to weights corresponding to normality in the simulation platform. Moreover, the weights of IA_1 would be initialized again when variations of parameters take place in the simulation platform. The system response and absolute value of the control error are presented in Figures 11(a) and 11(b), respectively. Intuitively, compared with the control mode used in Case 2, the control performance has improved dramatically. It is seen that the plant is controllable under different operating conditions. However, overshoots are somewhat large when parameters change in the plant. Particularly, at instant 90 s, a comparatively long period of time is needed for the adaptation process.

Case 3. Three range-based fixed models, one adaptive model, and one reinitialized adaptive model:

In previous sections, anomalous operating data have been mapped into 2-dimensional feature space via principle component analysis. Here, the establishment of $\{IF_i\}_{i=1}^3$ is accomplished by partitioning the range of first principle component (PC1). In this way, the threshold vector $\eta = [\eta_1, \eta_2]$ is obtained by evenly splitting the whole range of PC1 into three parts. Then, anomalous points corresponding to three intervals, namely, $[y_{\min}, \eta_1)$, $[\eta_1, \eta_2)$, and $[\eta_2, y_{\max}]$ ($y_{\min}$ and $y_{\max}$ are minima and maxima of PC1), are

FIGURE 10: Control response of Case 1: (a) system outputs; (b) absolute value of the control error.

FIGURE 11: Control response of Case 2: (a) system outputs; (b) absolute value of the control error.

collected and used to build three fixed models by MPDNN as before. IA_1 has the same setting as that in Case 2. Figure 12 presents the corresponding control response in this case. By comparing the absolute value of control error shown in Figure 12(b) with its counterpart Figure 11(b), it is found that the transient error at instant 30 s is reduced. Still, larger overshoot occurs when parameters change at 90 s.

Case 4. Three AMF clustering-based fixed models, one adaptive model, and one reinitialized adaptive model:

In Case 4, the determination of fixed models relies on the extracting results of the anomalous condition

obtained in Section 3. Therefore, three fixed models are constructed based on weights $\{\theta_i\}_{i=1}^3$. Meanwhile, the setting of IA_1 is consistent with the above cases. The resulting output and control error of the plant are depicted in Figures 13(a) and 13(b), respectively. It is clear that the control mode in this case can yield far better control performance in contrast to all cases mentioned above. From a transient response viewpoint, when parameters of the plant vary at different instants, the output can maintain small overshoot. On the contrary, the steady-state error can reach zero over time. The reason for such improvements lies in the reduction of initial parameter error in IA_2 . As revealed

FIGURE 12: Control response of Case 3: (a) system outputs; (b) absolute value of the control error.

FIGURE 13: Control response of Case 4: (a) system outputs; (b) absolute value of the control error; (c) switching sequence.

FIGURE 14: Control response of Case 4: (a) system outputs; (b) absolute value of the control error; (c) switching sequence.

from the switching sequence in Figure 13(c), each time when the type of anomalous condition shifts, the fixed model with the least performance index would be chosen to lower the transient error in the first place, and then IA_2 would take over as the controller until next environment appears. As the fixed models we built are able to roughly describe actual environments in different time ranges, transient errors are relatively small when switching to a new environment. Also, due to the closeness between parameters of fixed models and real ones, it takes little adaptation time for IA_2 to revert to set points.

Case 5. Additional fixed models on the basis of Case 4, one adaptive model, and one reinitialized adaptive model:

The design of Case 5 is built upon the last case. To further verify the effectiveness of fixed models

constructed in Case 4, additional three fixed models (IF_4, IF_5, IF_6) are considered. Further, weights of each additional fixed model are set randomly within the range of -1 and 1 . The system response and magnitude of error at every instant are presented in Figures 14(a) and 14(b) separately. Compared with their counterparts in Figure 13, it is clear that there is not much improvement in transient response with the variation of anomalous conditions. In fact, it can be seen that the system response in the normal operating condition is even worse. The reason can be concluded from the switching sequence shown in Figure 14(c). At the stage of controlling in normal condition, the parametric closeness of three additional fixed models to true parameters of the plant may give rise to occurrences of frequent switching, which causes intermittent overshoot between 60 and 90 units of time. Furthermore, during simulation, the amount of

computation became prohibitively large in this control mode, which proved once again that the inappropriate choice of model number when applying multimodel control can induce severe consequences in practical scenarios.

6. Conclusions

This paper focused on addressing the practical control problem in WLFGD system by using the MPDNN multimodel control strategy. The difficulties for the application of multimodel control in real world have been stated in Section 4. With the objective of implementing multimodel control in the context of anomalous conditions, a chain of techniques have been adopted to search and extract different types of anomalies. Then, MPDNN, which has an excellent capability of approximating uncertain environments, is utilized to identify those anomalous conditions and models are built accordingly. As indicated in Section 5, in comparison with other cases, the control performance of MPDNN multimodel control system, which is built upon the process we proposed in this work, can yield satisfactory control performance for simulated WLFGD system. Based on this direction, the main concern in our future work would be combinations of data analysis techniques and adaptive control in the context of multimodel control.

Data Availability

The experimental data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61873006 and 61673053) and National Key Research and Development Project (2018YFC1602704 and 2018YFB1702704).

References

- [1] T. Hlincik and P. Buryan, "Evaluation of limestones for the purposes of desulphurisation during the fluid combustion of brown coal," *Fuel*, vol. 104, pp. 208–215, 2013.
- [2] P. Fang, C. Cen, Z. Tang, P. Zhong, D. Chen, and Z. Chen, "Simultaneous removal of SO₂ and NO_x by wet scrubbing using urea solution," *Chemical Engineering Journal*, vol. 168, no. 1, pp. 52–59, 2011.
- [3] B. R. Deshwal, S. H. Lee, J. H. Jung, B. H. Shon, and H. K. Lee, "Study on the removal of NO_x from simulated flue gas using acidic NaClO₂ solution," *Journal of Environmental Sciences*, vol. 20, no. 1, pp. 33–38, 2008.
- [4] T. Lan, X. Zhang, Q. Yu, and L. Lei, "Study on the relationship between absorbed S(IV) and pH in the seawater flue gas desulfurization process," *Industrial & Engineering Chemistry Research*, vol. 51, no. 12, pp. 4478–4484, 2012.
- [5] T. Neveux and Y. Le Moullec, "Wet industrial flue gas desulfurization unit: model development and validation on industrial data," *Industrial & Engineering Chemistry Research*, vol. 50, no. 12, pp. 7579–7592, 2011.
- [6] S. Kiil, M. L. Michelsen, and K. Dam-Johansen, "Experimental investigation and modeling of a wet flue gas desulfurization pilot plant," *Industrial & Engineering Chemistry Research*, vol. 37, no. 7, pp. 2792–2806, 1998.
- [7] M. Gtomba, "Technical description of parameters influencing the pH value of suspension absorbent used in flue gas desulfurization systems," *Journal of the Air & Waste Management Association*, vol. 60, no. 8, pp. 1009–1016, 2010.
- [8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.
- [9] K. Severson, P. Chaiwatanodom, and R. D. Braatz, "Perspectives on process monitoring of industrial systems," *Annual Reviews in Control*, vol. 42, pp. 190–200, 2016.
- [10] R.-P. Nikula, E. Juuso, and K. Leiviskä, "Desulphurization plant monitoring and fault detection using principal component analysis," in *Proceedings of the 2013 8th EUROSIM Congress on Modelling and Simulation*, pp. 490–495, Cardiff, UK, September 2013.
- [11] Y.-p. Zheng and L.-p. Zhang, "Fault diagnosis of wet flue gas desulphurization system based on KPCA," in *Proceedings of the 19th International Conference on Industrial Engineering and Engineering Management*, pp. 279–288, Bangkok, Thailand, December 2013.
- [12] Z. Qiao, J. Zhou, F. Si, Z. Xu, and L. Zhang, "Fault diagnosis of slurry pH data base on autoregressive integrated moving average and least squares support vector machines," in *Proceedings of the 2013 Ninth International Conference on Natural Computation (ICNC)*, pp. 141–145, Shenyang, China, July 2013.
- [13] T. Korpela, T. Björkqvist, Y. Majanne, and P. Lautala, "Online monitoring of flue gas emissions in power plants having multiple fuels," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 1355–1360, 2014.
- [14] P. Córdoba, "Status of flue gas desulphurisation (FGD) systems from coal-fired power plants: overview of the physicochemical control processes of wet limestone FGDs," *Fuel*, vol. 144, pp. 274–286, 2015.
- [15] H. Gao, C. Li, G. Zeng et al., "Flue gas desulphurization based on limestone-gypsum with a novel wet-type PCF device," *Separation and Purification Technology*, vol. 76, no. 3, pp. 253–260, 2011.
- [16] R. Borsdorf, N. J. Higham, and M. Raydan, "Computing a nearest correlation matrix with factor structure," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2603–2622, 2010.
- [17] L. Puggini and S. McLoone, "An enhanced variable selection and isolation forest based methodology for anomaly detection with OES data," *Engineering Applications of Artificial Intelligence*, vol. 67, pp. 126–135, 2018.
- [18] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, Inc., Cambridge, MA, USA, 2009.
- [19] J. Handl and J. Knowles, "Feature subset selection in unsupervised learning via multiobjective optimization," *International Journal of Computational Intelligence Research*, vol. 2, no. 3, pp. 217–238, 2006.

- [20] I. Joliffe and B. Morgan, "Principal component analysis and exploratory factor analysis," *Statistical Methods in Medical Research*, vol. 1, no. 1, pp. 69–95, 1992.
- [21] S. Diana, *Principal Component Analysis vs. Exploratory Factor Analysis*, University of Northern Colorado, Greeley, CO, USA, 1999.
- [22] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, Pisa, Italy, December 2008.
- [23] X. Tao, Y. Peng, F. Zhao, P. Zhao, and Y. Wang, "A parallel algorithm for network traffic anomaly detection based on isolation forest," *International Journal of Distributed Sensor Networks*, vol. 14, no. 11, Article ID 1550147718814471, 2018.
- [24] L. Sun, S. Versteeg, S. Boztas, and A. Rao, "Detecting anomalous user behavior using an extended isolation forest algorithm: an enterprise case study," 2016, <http://arxiv.org/abs/1609.06676>.
- [25] S. Anand, S. Mittal, O. Tuzel, and P. Meer, "Semi-supervised kernel mean shift clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1201–1215, 2013.
- [26] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: a texture classification example," in *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV)*, pp. 456–463, Nice, France, October 2003.
- [27] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [28] D. Ming, T. Ci, H. Cai, L. Li, C. Qiao, and J. Du, "Semi-variogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 5, pp. 813–817, 2012.
- [29] X.-L. Li, C. Jia, K. Wang, and J. Wang, "Trajectory tracking of nonlinear system using multiple series-parallel dynamic neural networks," *Neurocomputing*, vol. 168, pp. 1–12, 2015.
- [30] B. Widrow and S. D. Steans, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.
- [31] A. S. Poznyak, W. Wen Yu, and E. N. Sanchez, "Identification and control of unknown chaotic systems via dynamic neural networks," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 46, no. 12, pp. 1491–1495, 1999.
- [32] A. S. Poznyak, W. Wen Yu, E. N. Sanchez, and J. P. Perez, "Nonlinear adaptive trajectory tracking using dynamic neural networks," *IEEE Transactions on Neural Networks*, vol. 10, no. 6, pp. 1402–1411, 1999.
- [33] W. Yu and A. S. Poznyak, "Indirect adaptive control via parallel dynamic neural networks," *IEEE Proceedings—Control Theory and Applications*, vol. 146, no. 1, pp. 25–30, 1999.
- [34] J. Willems, "Least squares stationary optimal control and the algebraic Riccati equation," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 621–634, 1971.
- [35] W. Yu, A. S. Poznyak, and E. N. Sanchez, "Dynamic multi-layer neural networks for nonlinear system on-line identification," in *Proceedings of the 2000 IEEE International Symposium on Intelligent Control. Held Jointly with the 8th IEEE Mediterranean Conference on Control and Automation (Cat. No. 00CH37147)*, pp. 25–30, Rio Patras, Greece, July 2000.
- [36] K. S. Narendra, J. Balakrishnan, and M. K. Ciliz, "Adaptation and learning using multiple models, switching, and tuning," *IEEE Control Systems Magazine*, vol. 15, no. 3, pp. 37–51, 1995.
- [37] K. S. Narendra, O. A. Driollet, M. Feiler, and K. George, "Adaptive control using multiple models, switching and tuning," *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 2, pp. 87–102, 2003.
- [38] B. Dou, W. Pan, Q. Jin, W. Wang, and Y. Li, "Prediction of SO₂ removal efficiency for wet flue gas desulfurization," *Energy Conversion and Management*, vol. 50, no. 10, pp. 2547–2553, 2009.
- [39] J. Li, X. Guo, Z. Li, and W. Chen, "Stochastic adaptive optimal control of under-actuated robots using neural networks," *Neurocomputing*, vol. 142, pp. 190–200, 2014.
- [40] R. Piziak and P. L. Odell, "Matrix theory. From generalized inverses to Jordan form," *Bulletin of the American Mathematical Society*, vol. 30, pp. 417–453, 2007.