Journal of Southeast University (English Edition) Vol. 38, No. 4, pp. 383 - 391 Dec. 2022 ISSN 1003—7985

Dynamic model for predicting nitrogen oxide concentration at outlet of selective catalytic reduction denitrification system based on kernel extreme learning machine

Ma Ning[1] Liu Lei[1] Yang Zhenyong[1] Yan Laiqing[2] Dong Ze[3]

([1]North China Electric Power Research Institute Co., Ltd., Beijing 100045, China)

([2]School of Electric Power, Civil Engineering and Architecture, Shanxi University, Taiyuan 030006, China)

([3]Hebei Technology Innovation Center of Simulation and Optimized Control for Power Generation, North China Electric Power University, Baoding 071003, China)

Abstract: To solve the increasing model complexity due to several input variables and large correlations under variable load conditions, a dynamic modeling method combining a kernel extreme learning machine (KELM) and principal component analysis (PCA) was proposed and applied to the prediction of nitrogen oxide (NO_x) concentration at the outlet of a selective catalytic reduction (SCR) denitrification system. First, PCA is applied to the feature information extraction of input data, and the current and previous sequence values of the extracted information are used as the inputs of the KELM model to reflect the dynamic characteristics of the NO_x concentration at the SCR outlet. Then, the model takes the historical data of the NO_x concentration at the SCR outlet as the model input to improve its accuracy. Finally, an optimization algorithm is used to determine the optimal parameters of the model. Compared with the Gaussian process regression, long short-term memory, and convolutional neural network models, the prediction errors are reduced by approximately 78.4%, 67.6%, and 59.3%, respectively. The results indicate that the proposed dynamic model structure is reliable and can accurately predict NO_x concentrations at the outlet of the SCR system.

Key words: selective catalytic reduction; nitrogen oxides; principal component analysis; kernel extreme learning machine; dynamic model

DOI: 10.3969/j.issn.1003-7985.2022.04.007

In recent years, with the pursuit of a higher quality of life and deterioration of air quality, the Chinese government has become more stringent regarding the emission requirements of nitrogen oxide (NO_x)[1]. Selective catalytic reduction (SCR) flue gas denitrification is a technique applied in coal-fired power plants to reduce NO_x. The SCR denitrification technique has some salient features, including a high denitrification efficiency and simple device structure, and has thus attracted significant attention and a wide application in most power plants in China[2-3].

However, based on practical experience, the efficiency of the SCR denitrification system is easily affected by the amount of ammonia injection and diluted air, reaction temperature, catalyst activity, and other factors. When the design and finalization of a unit are is put into operation, the reaction temperature can be controlled via a flue gas bypass, and the catalyst can be replaced after failure. Thus, the amount of ammonia injection is the key factor for a daily adjustment in the control of NO_x emissions[4]. In addition, an increase in the amount of escaping ammonia will increase the operation cost and secondary pollution in the environment[5]. Thus, a suitable amount of ammonia injection is essential for the operation of an SCR denitrification system[6].

Establishing an accurate model for predicting NO_x concentrations at the SCR system outlet is a prerequisite for the implementation of an accurate ammonia injection. The NO_x concentration at the outlet of the SCR is influenced by several thermal parameters, such as the entrance NO_x concentration, inlet gas flow value, ammonia injection, and unit load. A strong nonlinearity, coupling, and inertia occur between the factors and the NO_x concentration at the SCR outlet. These characteristics increase the difficulty regarding the development of a mechanistic model when applying a first-principle analysis method.

Fortunately, a data-driven modeling method as an efficient modeling method has been widely applied to various industrial fields[7]. The extreme learning machine (ELM) technique was proposed by Huang et al.[8] Compared with a traditional backpropagation (BP) neural network or radial basis function (RBF) network, ELM can avoid overfitting and has a relatively fast learning speed. However, the performance of an ELM is affected to a great extent by the number of nodes in its hidden layer. To deal with this issue, Huang et al.[9] proposed the use of a kernel extreme learning machine (KELM), which is

Received 2022-05-29, **Revised** 2022-08-30.

Biography: Ma Ning (1992—), male, doctor, engineer, maningncepu@163.com.

Foundation items: The National Natural Science Foundation of China (No. 71471060), the Natural Science Foundation of Hebei Province (No. E2018502111).

Citation: Ma Ning, Liu Lei, Yang Zhenyong, et al. Dynamic model for predicting nitrogen oxide concentration at outlet of selective catalytic reduction denitrification system based on kernel extreme learning machine [J]. Journal of Southeast University (English Edition), 2022, 38(4): 383 - 391. DOI: 10.3969/j.issn.1003 - 7985.2022.04.007.

an extension of an ELM method.

Static models are also known as steady-state models and are generally developed using data under steady-state conditions to achieve the goal of optimization or monitoring of a power plant[10]. However, the introduction of an automatic generation control system can result in frequent changes in the operational parameters of a power plant. The SCR denitrification system has strong nonlinearity and inertia owing to its working principle. Hence, common static models are inadequate for accurately reflecting the characteristics of NO_x concentrations at the SCR system outlet. Compared to a steady-state model, it is more difficult to establish a dynamic model of the thermal process using a data-driven method[11].

In this study, the KELM method is applied to establish a model that describes the dynamic characteristics of the NO_x concentration at the SCR denitrification system outlet of a 1 000 MW ultra-supercritical unit. First, the principal component analysis (PCA) method is used to extract the characteristic information from the initial input data to reduce the correlation between the input data and input dimension. The proposed dynamic model uses the PCA method to extract the input data features. Its advantages are reflected in two aspects. On the one hand, PCA reduces the correlation of input variables, which consequently reduces the correlation coupling between the final model inputs and is conducive to the establishment of a data model with high prediction accuracy. On the other hand, the number of input variables can be reduced through feature extraction. Next, the current and previous sequence values of the extracted information are used as the inputs of the KELM model, and the parameters of the KELM are optimized through quantum particle swarm optimization (QPSO). Finally, the dynamic KELM model is developed using the actual data of a power plant operation.

1 Theory and Algorithm

1.1 KELM

The ELM is a feedforward neural network with a single hidden layer. The bias and weights of the input layer of the ELM algorithm are randomly and independently assigned. Given that N training samples $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $\boldsymbol{x}_i \in \mathbf{R}^n$ represents the input data, and $\boldsymbol{y}_i \in \mathbf{R}^m$ is the output data. In addition, n and m represent the dimensions of the input and output data, respectively. The standard mathematical model of the ELM can be expressed as follows[8]:

$$\sum_{i=1}^{l} \boldsymbol{\beta}_i g_i(\boldsymbol{x}_j) = \sum_{i=1}^{l} \boldsymbol{\beta}_i g(\boldsymbol{w}_i \boldsymbol{x}_j + \boldsymbol{b}_i) = \boldsymbol{o}_j \quad j = 1, 2, \cdots, N \tag{1}$$

where $\boldsymbol{\beta}_i$ is the output weight, which connects the i-th hidden node with the output nodes; $\boldsymbol{w}_i$ represents the input weights connecting the i-th hidden node with the input nodes. Previous studies have shown that the output value of the ELM model can be fitted to the samples with zero error. Thus, a derivation equation can be obtained as follows, where the number of the hidden layer nodes of the model is l, and the activation function is $g(\cdot)$:

$$\sum_{j=1}^{l} \| \boldsymbol{o}_j - \boldsymbol{y}_j \| = 0 \qquad j = 1, 2, \cdots, N \tag{2}$$

The above equation can be simply written as

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{y} \tag{3}$$

where

$$\boldsymbol{H} = \begin{bmatrix} g(\boldsymbol{w}_1 \boldsymbol{x}_1 + b_1) & g(\boldsymbol{w}_2 \boldsymbol{x}_2 + \boldsymbol{b}_2) & \cdots & g(\boldsymbol{w}_l \boldsymbol{x}_1 + \boldsymbol{b}_l) \\ \vdots & \vdots & & \vdots \\ g(\boldsymbol{w}_1 \boldsymbol{x}_N + \boldsymbol{b}_1) & g(\boldsymbol{w}_2 \boldsymbol{x}_N + \boldsymbol{b}_2) & \cdots & g(\boldsymbol{w}_l \boldsymbol{x}_N + \boldsymbol{b}_l) \end{bmatrix}_{N \times l} \tag{4}$$

$$\boldsymbol{\beta} = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_1\}^{\mathrm{T}}, \ \boldsymbol{y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N\}^{\mathrm{T}} \tag{5}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{y}$ are the output weight matrix and output data matrix, respectively.

The output weight $\boldsymbol{\beta}$ can be calculated using the following equation:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{H}^{+} \boldsymbol{y} \tag{6}$$

where $\boldsymbol{H}^{+}$ is the Moore-Penrose generalized inverse of $\boldsymbol{H}$.

To increase the robustness and generalization capability of the ELM, a coefficient C and kernel were proposed by Huang et al. based on an analysis of the support vector machine theory. The KELM uses Mercer's conditions to define the kernel matrix $\boldsymbol{\Omega}$:

$$\boldsymbol{\Omega} = \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}; \boldsymbol{\Omega} = h(\boldsymbol{x}_i) h(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{7}$$

Based on the above equations, the output of the KELM is determined as follows[9]:

$$f(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}) \boldsymbol{H}^{\mathrm{T}} \left(\frac{1}{C}\boldsymbol{I} + \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}} \right)^{-1} \boldsymbol{y} = \begin{bmatrix} K(\boldsymbol{x}, \boldsymbol{x}_1) \\ K(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ \vdots \\ K(\boldsymbol{x}, \boldsymbol{x}_N) \end{bmatrix}^{\mathrm{T}} \left(\frac{1}{C}\boldsymbol{I} + \boldsymbol{\Omega} \right)^{-1} \boldsymbol{y} \tag{8}$$

In this study, the RBF is selected as the kernel function:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(- \frac{\| \boldsymbol{x}_i - \boldsymbol{x}_j \|^2}{\gamma^2} \right) \tag{9}$$

where γ represents the kernel parameter of the KELM.

1.2 PCA

Suppose $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m\}$, with m variables and n samples. The matrix $\boldsymbol{x}$ can be expressed using the PCA as

follows[12]:

$$\boldsymbol{t}_i = \boldsymbol{a}_i^{\mathrm{T}} x \tag{10}$$

where $\boldsymbol{a}$ is the orthogonal matrix, which can be calculated as

$$|\boldsymbol{R} - \lambda \boldsymbol{I}| = 0 \tag{11}$$

where $\boldsymbol{R}$ and $\boldsymbol{l}$ indicate Pearson's coefficients of matrix $\boldsymbol{x}$ and $\boldsymbol{a}$ can be calculated using the corresponding eigenvectors. In addition, $\boldsymbol{t}_i$ is a newly extracted principal component. To qualitatively describe the explanation rate provided by the principal component, the variance contribution rate δ_i of the principal component $\boldsymbol{t}_i$ and the cumulative explanation rate of the former p principal component h_p are defined as

$$\delta_i = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j}, \quad \eta_p = \sum_{i=1}^{p} \delta_i = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \tag{12}$$

The PCA achieves a reduction in the dimension by eliminating strictly linear or highly correlated independent variable information while ignoring principal components with low explanation rates.

2 SCR Denitrification System and Data Preparation

This study mainly considers the SCR flue gas denitrification system of a 1 000 MW ultra-supercritical unit in the Tai Zhou power plant in Jiangsu Province, China. Fig. 1 shows the general layout of the SCR flue gas denitrification system. SCR devices are installed between the air preheater and the economizer of the boiler[13]. TiO_2 is used as a catalyst in this SCR ammonia injection flue gas denitrification system. The diluted air is mixed with ammonia from the ammonia station. The ammonia injection flow rate is adjusted using an ammonia injection control valve. Then, the diluted air is ejected through the nozzle, and the flue gas is fully mixed. Under the catalysis of the TiO_2, mixed gases can be selectively catalyzed to reduce into harmless N_2 and H_2O, thus achieving the purpose of flue gas denitrification. The main chemical reaction equation of SCR denitrification is as follows[13]:

$$\left.\begin{aligned} 4NH_3 + 4NO + O_2 &= 4N_2 + 6H_2O \\ 4NH_3 + 6NO &= 5N_2 + 6H_2O \end{aligned}\right\} \tag{13}$$

A high-quality modeling sample is essential for building an accurate data-driven model. In the Tai Zhou power plant, significant amounts of previous operational data are continuously stored in a distributed control system database using various acquisition equipment. In this study, the boiler load in the data segment varies from 700 to 1 000 MW. In addition, based on basic knowledge of boilers and engineering experience[14], six variables are employed as inputs of the SCR model. The only output is the export of the NO_x concentration of the SCR denitrification system. A description of the variables and parameter ranges is listed in Tab. 1. There is a significant difference between the dynamic model and the steady-state model in the selection of the modeling samples. Steady-state models tend to select data that are mostly invariable or with little variation in the operating parameters, which can be selected from different time periods. By contrast, data samples used to build dynamic models should be continuous, and data segments can reflect the characteristics of the changes in the production process. Based on the above-mentioned criteria for a dynamic model sample selection, a total of 13 000 operational data samples are obtained, and the time sampling interval is 1 min. The data samples are shown in Fig. 2, where a variation of the unit load can be clearly observed. In this study, 4 200 measurements of boiler operation are selected as the training samples to train the KELM model. Another set of 3 700 data samples is used as the testing samples to evaluate

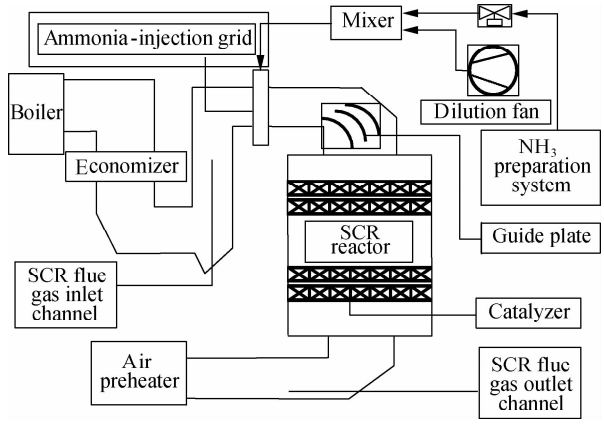


Fig. 1 Schematic diagram of the reactor structure in the SCR flue gas denitrification system

Tab. 1 Description of variables and parameter ranges

Variable description	Variable range
Entrance NO_x concentration/(mg · m^{-3})	121.94 to 279.96
Inlet gas flow value/(m^3 · h^{-1})	131.56 to 171.00
Inlet flue gas temperature/℃	353.99 to 378.13
Ammonia injection/(kg · h^{-1})	46.52 to 105.73
Boiler load/MW	718.81 to 981.28
Entrance O_2 concentration/(mg · m^{-3})	3.21 to 5.61
Export NO_x concentration/(mg · m^{-3})	25.32 to 58.26

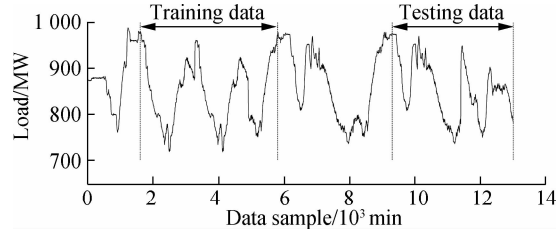


Fig. 2 Load range chart of the selected sample data

the generalized performance of the model.

Data preprocessing is an indispensable step in establishing a data-driven model. All unusual operation data samples and outliers that fail to correctly respond to the process characteristics due to a sensor failure or equipment abnormality should be removed. Many methods for detecting the outliers of the modeling samples have been proposed. The core idea of these methods is to eliminate samples that are clearly different from nearby data and then use specific technical means to complete deleted data and meet data continuity. Data standardization can eliminate the impact of different variable scales. The input and output data are preprocessed in the same order as the amplitude using the following equations:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad y' = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \tag{14}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of input variables x, respectively; x' is the standardized input variable. Similarly, $y_{\max}$ and $y_{\min}$ are the maximum and minimum values of output variables y, respectively; y' is the standardized output variable.

3 Development of the Dynamic SCR Denitrification Model

3.1 Model input feature extraction using PCA

For the six input variables selected to establish the NO_x concentration model for the SCR outlet, a certain degree of correlation exists among the variables. For example, there is a significant correlation between the unit load and inlet flue gas flow because an increase in the unit load will inevitably accompany an increase in the boiler fuel and total. As a result, the flue gas flow of the SCR denitrification system will also increase. If these variables are directly used as the model input, a redundancy of the input information will inevitably weaken the generalization capability of the model. To eliminate the correlation among the variables, the PCA technique is used to extract information about the input variables before building the model.

According to Eq. (12), the interpretation rate of the information variance is determined based on the number of extracted principal components. The explanation rate of the six principal components for the input variable data used to establish the SCR model is illustrated in Fig. 3. The explanation rate of the first two principal components is as much as 95% of the primary data samples and is obviously higher than that of the other principal components. Hence, the original six input variables have a large degree of linear dependence on one another, which is also consistent with the actual operation. When the number of principal components is 2, approximately 95% of the variance information in the dependent variables is explained. At this time, the information regarding the variance interpretation obtained by extracting new principal components slightly changes, and thus the residual can be considered noise interference. If the number of principal components is increased, the noise will be introduced and the complexity of the model increases, which can reduce the prediction accuracy of the model. Hence, the first two principal components are applied as input vectors for establishing the KELM model.

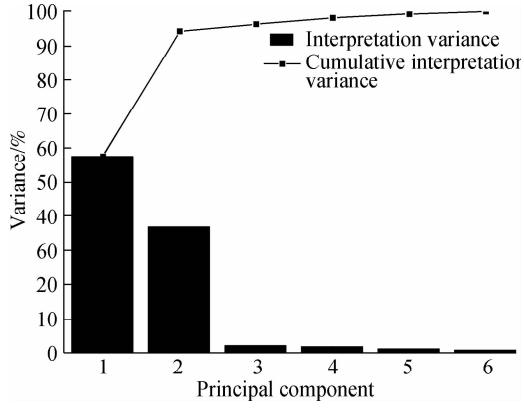


Fig. 3 Explanation rate of the six principal components and cumulative explanation rate

3.2 Dynamic model structure for SCR denitrification

When setting up a steady-state model of the industrial process, the data samples of the independent variables and predicted variables must be in a stable state. In other words, the core idea of establishing a steady model is to find a functional expression between the current values of the input and output variables at the current time using a mathematical method. Suppose the KELM is used to establish a steady-state model; the prediction model can be expressed as follows:

$$y(t) = f(\boldsymbol{x}(t)) \tag{15}$$

where $y(t)$ and $\boldsymbol{x}(t)$ are the predicted and independent variables at time t, respectively; $f(\cdot)$ is the mapping function of the KELM.

However, an SCR denitrification system has a large delay and strong dynamic characteristics. Considering that the time sequence of the input and output data is the most significant difference between static and dynamic models if only the current time value of the independent variable is considered the input of the model without considering the previous period of time; the performance of the model in reflecting the dynamic characteristics of the process will be restricted. The past values of the output can also influence the current output value of the dynamic system. Hence, input and output variable delays at different scales are often introduced into dynamic models. For a given input variable $\boldsymbol{x}$ and output variable y, $(\boldsymbol{x}(t), y(t))$ represent the input and output values of the sample at time

t. The inputs of a dynamic model should be composed of $\boldsymbol{x}(t)$ and previous time values $\boldsymbol{x}(t-\tau)$, $y(t-\tau_y)$. The dynamic prediction model can be expressed as follows:

$$\begin{aligned} y(t) &= f(\boldsymbol{X}(t)) \\ \boldsymbol{X}(t) &= \{x_1(t), x_1(t-1), \cdots, x_1(t-\tau_1), \cdots, x_m(t), \\ & x_m(t-1), \cdots, x_m(t-\tau_m), \cdots, y(t-1), \cdots, y(t-\tau_y)\} \end{aligned} \tag{16}$$

where m is the number of dimensions of the independent variable $\boldsymbol{x}$; τ_m and τ_y are the delay order sizes of $\boldsymbol{x}$ and y, respectively.

Based on the principle of a dynamic model and the PCA, the structure of the SCR denitrification dynamic model is shown in Fig. 4. Two principal components, u_1 and u_2, are extracted from the original input data using the PCA method. For u_1 and u_2, two time series, whose ranges of variation vary from the current time value $u_1(t)$ to the past time $u_1(t-\tau_1)$ and from $u_2(t)$ to $u_2(t-\tau_2)$, are considered inputs of the model. In addition, the past values of the predicted variable y are applied as the model inputs. Accordingly, the new input dimensions of the dynamic model are $d = 1 + \tau_1 + 1 + \tau_2 + \tau_y$.

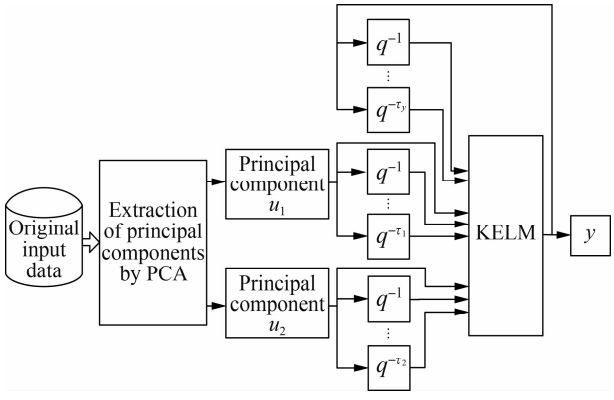


Fig. 4 Structure of the dynamic model used in SCR denitrification

3.3 Parameter selection method of the SCR dynamic model

The performance of the KELM model is usually affected by the selection of the kernel parameters γ and regularization coefficient C. The two parameters can be determined through k-fold cross-validation during the experiments. Specifically, the training samples are equally divided into k groups. Then, the $k-1$ groups are used to train the KELM model, and the remaining group is applied to test the model. After k repeated experiments, each group of data can be used as test data in turn. The average of the total test errors is taken as an assessment criterion to evaluate the parameters of the KELM model. In this study, the 10-fold cross-validation method and QPSO are used to optimize the combination of parameters (γ, C). The search range of the kernel parameter γ is $(0, 500)$, and the search range of the regularization coefficient C is $(0, 600)$. The QPSO algorithm is an intelligent optimization algorithm based on the particle swarm optimization (PSO) algorithm and quantum mechanics theory. The QPSO algorithm overcomes the shortcomings of a low aggregation performance and limited search range in the PSO algorithm to a certain extent and has a good particle search performance. In recent years, the QPSO algorithm has been widely used for solving difficult optimization problems in numerous fields[15]. In this work, the QPSO algorithm parameters are set as follows: the search particle population is 30, the maximum number of iterations is 200, and the initial search particle position is the random position within the search interval. In addition, the delay orders of the dynamic models, including τ_1, τ_2, and τ_y, need to be determined during the model development. Owing to the lack of a theoretical basis, selecting the appropriate delay order of each parameter is more difficult than determining the kernel parameters and regularization coefficients. Although an open-loop test is a theoretically feasible method, it requires a lot of money and on-site resources. In the present study, the delay orders of dynamic models are calculated using a trial-and-error approach. The detailed process is given in the following section.

4 Results and Discussions

To quantitatively evaluate the performance of the proposed dynamic models of the SCR denitrification system, the root mean square error (RMSE) and mean relative error (MRE) are applied as the evaluation criteria, which can be written as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}},\ \mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{17}$$

where n represents the number of samples and y_i and $\hat{y}_i$ are the real and corresponding predicted values, respectively.

To prove the rationality of the model structure, several models with different numbers of input delays are compared. The first model we established is denoted as model Ⅰ, where delays in the input and output variables are not considered. In other words, the input variables of model Ⅰ only include the extracted principal component $u_1(t)$ and principal component $u_2(t)$. Model Ⅰ is a steady-state model of an SCR denitrification system in a coal-fired power plant. The RMSE and MRE of model Ⅰ for the training samples are 3.306 3 mg/m^3 and 6.85%, respectively. For the testing samples, the ξ_{RMSE} and ξ_{MRE} of model Ⅰ are 4.794 0 mg/m^3 and 10.35%, respectively. In addition, the prediction results of the testing samples

are shown in Fig. 5, where the error in the predicted results for the testing samples is large, and that of the individual special sample points even exceeds 20 mg/m^3. Hence, for data with strong dynamic characteristics, the steady-state model without considering the delay cannot achieve good prediction results.

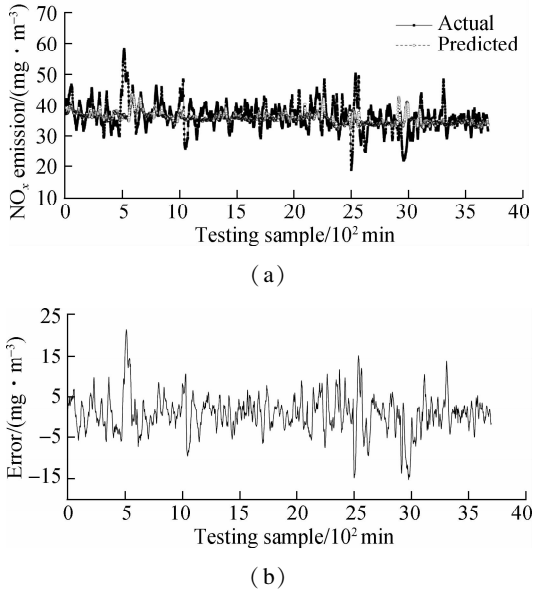


Fig. 5 Predicted performance of model Ⅰ for testing data. (a) Fitting curve; (b) Prediction error

Model Ⅱ is established by considering the delay of the NO_x concentration at the SCR outlet y. In contrast to model Ⅰ, the input variables of model Ⅱ contain the historical sequence of y, and its mathematical expression can be written as follows:

$$y(t) = f(\boldsymbol{X}(t))$$
$$\boldsymbol{X}(t) = \{u_1(t), u_2(t), y(t-1), \cdots, y(t-\tau_y)\} \tag{18}$$

When τ_y equals zero, the delay of output y is not considered, and model Ⅱ becomes the same steady-state model as model Ⅰ. Fig. 6 shows the variation in the ξ_{RMSE} values of the training samples and testing data with a τ_y value from 0 to 10. When τ_y is assigned as 2, model Ⅱ achieves the minimum ξ_{RMSE} value of the testing sample, which is 0.803 5 mg/m^3. Compared with model Ⅰ, the prediction accuracy of model 2 is improved by 83.23%, which indicates that the generalization ability of the model can be greatly improved by introducing the output delay τ_y into the model input.

To evaluate the effect of the delay in the extracted component u_1 on the model performance, we set up model Ⅲ, where the historical sequence of u_1 is added to the model inputs. The inputs of model Ⅲ can be written as $\boldsymbol{x}(t) = \{u_1(t), u_1(t-1), \cdots, u_1(t-\tau_1), u_2(t)\}$. The variation in the ξ_{RMSE} values of model Ⅲ for testing the sample data with the τ_1 value from 0 to 10 is shown in

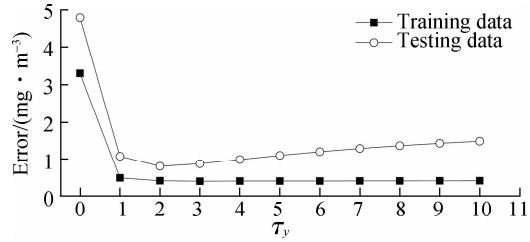


Fig. 6 Variation of the RMSE values with the output delay τ_y

Fig. 7. In the figure, the prediction accuracy of the model gradually decreases and then increases with an increase in τ_1. The minimum error is obtained when τ_1 is set to 3 with ξ_{RMSE} = 4.657 5 mg/m^3. The prediction accuracy of model Ⅲ is slightly higher than that of model Ⅰ but not as high as that of model Ⅱ. Hence, adding the historical data of the principal component u_1 to the model input is beneficial for improving the prediction performance.

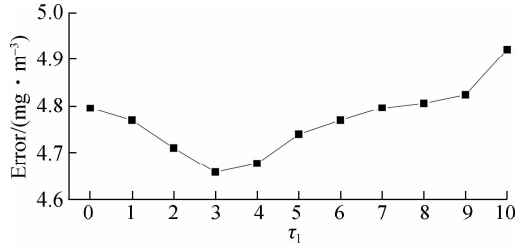


Fig. 7 Variation in RMSE values with τ_1

With model Ⅲ, to assess the effect of the delay in the extracted component u_2 on the model performance, model Ⅳ is built. Its input can be expressed as $\boldsymbol{x}(t) = \{u_1(t), u_2(t), u_2(t-1), \cdots, u_2(t-\tau_2), y(t-1), \cdots, y(t-\tau_y)\}$. Figs. 8(a) and (b) show the prediction results of model Ⅳ for the testing samples, in which τ_y is set to 0 and 2, respectively. In Fig. 8(a), where τ_y is set to 0,

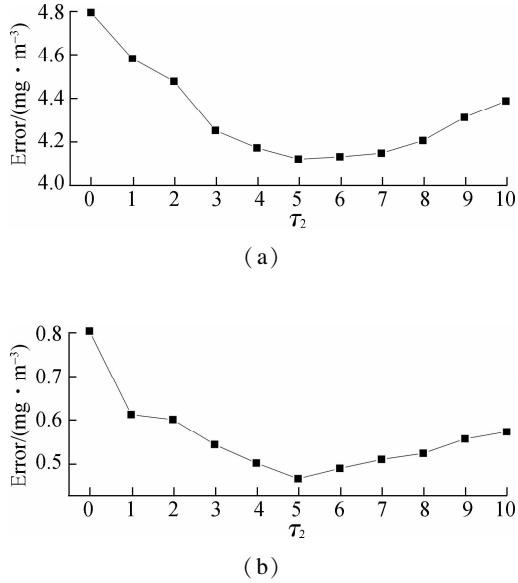


Fig. 8 Variation of RMSE values with τ_2. (a) τ_y = 0; (b) τ_y = 2

i.e., the delay of output y is not considered, the minimum error of the model is reached when τ_2 is increased to 5 and ξ_{RMSE} is 4.120 1 mg/m^3. Another case is shown in Fig. 8(b), where τ_y is set to 2, and the optimal error value of 0.465 5 mg/m^3 is obtained when τ_2 is set to 5. In both cases, the generalization performance of the model is improved when compared with that of the model in which the extracted component u_2 delay is not added to the model inputs.

By introducing the historical data of the original input variables u_1 and u_2 and output y into the model inputs, the improvement in the prediction accuracy of the model can be seen at various degrees. To further improve the prediction accuracy of the dynamic model, model Ⅴ is developed by adding the delays of u_1, u_2, and y. The inputs of model Ⅴ are expressed as $\boldsymbol{x}(t) = \{u_1(t), u_1(t-1), \cdots, u_1(t-\tau_1), u_2(t), u_2(t-1), \cdots, u_2(t-t_2), y(t-1), \cdots, y(t-\tau_y)\}$. A grid search method is applied to determine the delay values of τ_1, τ_2, and τ_y. Based on the relationship between the delay and accuracy of the model, the search range of the parameters is set as follows: the ranges of τ_1 and τ_2 are set to [1, 10], and the range of τ_y is set to [1, 5]. Fig. 9 shows the grid search results when τ_y is fixed at 2. By optimizing the search, the optimum parameters are determined as $\tau_1 = 4$, $\tau_2 = 5$, and $\tau_y = 2$. The minimum ξ_{RMSE} of the testing data is 0.422 9 mg/m^3. The corresponding prediction result is shown in Fig. 10, where the predicted curve of model Ⅴ mostly coincides with the actual curve. In addition, Tab. 2 shows the parameter information and the error accuracy on the training and testing samples of the five different models. Compared with the first four models, the training and generalization performance of model Ⅴ are significantly improved. This result indicates that introducing appropriate historical data into the model input can greatly enhance the capability of the model in describing the dynamic characteristics of the NO_x concentration at the outlet of the SCR denitrification system.

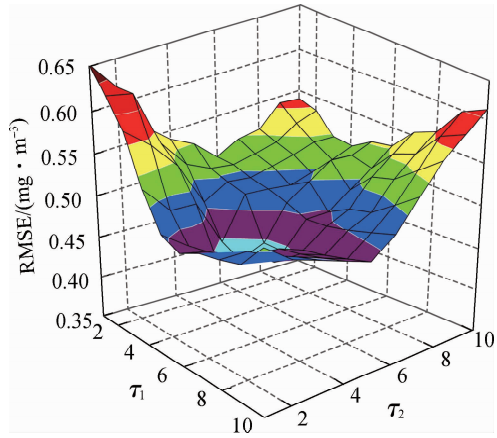
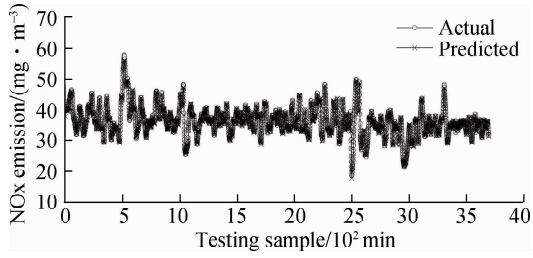
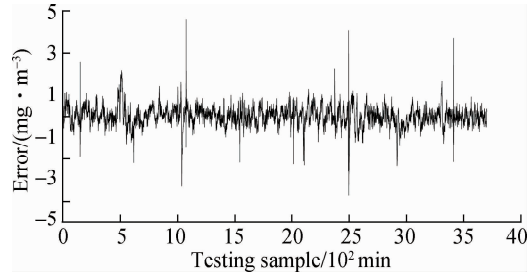


Fig. 9 Variation of the RMSE values with τ_1 and τ_2



(a)



(b)

Fig. 10 Predicted performance of model Ⅴ for testing data. (a) Fitting curve; (b) Prediction error

Tab. 2 Parameter setting and prediction results of the five models

Model	Parameter			Training sample		Testing sample		γ	C
	τ_1	τ_2	τ_y	RMSE/(mg · m^{-3})	RME/%	RMSE/(mg · m^{-3})	RME/%		
Ⅰ	0	0	0	3.306 3	6.85	4.794 0	10.35	81.56	0.21
Ⅱ	0	0	2	0.448 7	0.86	0.803 5	1.47	142.84	45.80
Ⅲ	3	0	0	3.005 3	6.22	4.657 5	9.85	100.32	96.84
Ⅳ	0	5	0	2.362 7	4.90	4.120 1	8.30	168.12	25.26
	0	5	2	0.403 7	0.77	0.465 5	0.89	121.22	85.25
Ⅴ	4	5	2	0.375 3	0.71	0.422 9	0.84	154.32	78.45

To further verify the performance of the proposed model, three other modeling methods are used for experimental comparison. The three methods are Gaussian process regression (GPR)[16], long short-term memory (LSTM) neural network[17], and convolutional neural network (CNN)[18]. The results of the four models are displayed in Tab. 3. In the table, for the GPR approach, the ξ_{RMSE} value of the GPR model is 1.031 1 for the training dataset and 1.989 3 for the testing dataset. The LSTM and CNN methods display an obvious betterment, and their root mean square errors of the test data are 1.307 8 and 1.040 7, respectively. Compared with the GPR, LSTM, and CNN models, the prediction errors are reduced by approximately 78.4%, 67.6%, and 59.3%, respectively. As the

four models use the same training and testing datasets, the predicted results on the testing dataset are reliable in proving the models' generalization and prediction accuracy capability. The GPR model has the worst performance among the four models. The LSTM and CNN perform better than GPR but are not as good as model Ⅴ. Therefore, the proposed dynamic model is more precise than the three other models for modeling the NO_x concentration at the outlet of the SCR denitrification system.

Tab. 3 Comparison of the prediction results of the four models

Model	Training samples		Testing samples	
	RMSE/(mg·m^{-3})	RME/%	RMSE/(mg·m^{-3})	RME/%
Model Ⅴ	0.375 3	0.71	0.422 9	0.84
GPR	1.031 1	1.74	1.989 3	2.80
LSTM	0.873 3	1.47	1.307 8	1.83
CNN	0.425 7	0.81	1.040 7	1.52

5 Conclusions

1) A dynamic model for the NO_x concentration at the outlet of an SCR denitrification system is established based on a combination of the KELM and PCA methods. The current and previous values of the two principal components extracted from the original input data are applied as model inputs. In addition, the value of the NO_x concentration at the outlet of the SCR denitrification system for a period of time is introduced into the model inputs as feedback.

2) The performance of the dynamic model is validated by the real operational data of a 1 000 MW ultra-supercritical unit. Compared with the GPR, LSTM, and CNN models, the prediction errors were reduced by approximately 78.4%, 67.6%, and 59.3%, respectively.

3) The comparison results demonstrate that the proposed model has a relatively high precision prediction performance and can be an alternative when designing a software package to control NO_x emissions from the SCR denitrification system.

References

[1] Yang G, Wang Y, Li X. Prediction of the NO_x emissions from thermal power plant using long-short term memory neural network[J]. *Energy*, 2020, **192**:116597. DOI: 10.1016/j.energy.2019.116597.

[2] Kang J, Niu Y, Hu B. Dynamic modeling of SCR denitration systems in coal-fired power plants based on a Bi-directional long short-term memory method[J]. *Process Saf Environ Prot*, 2021, **192**: 867-878. DOI: 10.1016/j.psep.2021.02.009.

[3] Fu J, Xiao H, Wang H, et al. Control strategy for denitrification efficiency of coal-fired power plant based on deep reinforcement learning[J]. *IEEE Access*, 2020, **8**: 65127-65136. DOI: 10.1109/ACCESS.2020.2985233.

[4] Dong Z, Ma N. A novel nonlinear partial least square integrated with error-based extreme learning machine[J]. *IEEE Access*, 2019, **7**: 59903-59912. DOI: 10.1109/ACCESS.2019.2911741.

[5] Wu X, Shen J, Li Y et al. Steam power plant configuration, design and control[J]. *Wiley Interdiscip Rev: Energy Environ*, 2015, **4**: 537-563.

[6] Niu Y, Pan Y, Huang W. Artificial fish swarm and feedback linearization of flue gas denitration control based on neural network[J]. *Journal of System Simulation*, 2018, **30**:2707-2714. (in Chinese)

[7] Pham B, Shirzadi A, Bui D, et al. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area [J]. *India Int J Sediment Res*, 2018, **2**: 157-170. DOI: 10.1016/j.ijsrc.2017.09.008.

[8] Huang G, Zhu Q, Siew C. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, **70**: 489-501. DOI: 10.1016/j.neucom.2005.12.126.

[9] Huang G, Wang D, Lan Y. Extreme learning machines: a survey[J]. *Int J Mach Learn Cybern*, 2011, **2**: 107-122. DOI: 10.1007/s13042-011-0019-y.

[10] Dong Z, Ma N, Meng L. Model improvement for boiler NO_x emission based on DEQPSO algorithm[J]. *Journal of Chinese Society of Power Engineering*, 2019, **39**: 191-197. (in Chinese)

[11] Kadlec P, Grbic R, Gabrys B. Review of adaptation mechanisms for data-driven soft sensors[J]. *Comput Chem Eng*, 2011, **35**: 1-24.

[12] Geladi P. Notes on the history and nature of partial least squares (PLS) modeling[J]. *J Chemom*, 1988, **2**: 231-246. DOI: 10.1002/cem.1180020403.

[13] Sjövall H, Blint R, Gopinath, A. A kinetic model for the selective catalytic reduction of NO_x with NH_3 over an Fe-zeolite catalyst[J]. *Ind Eng Chem Res*, 2010, **49**: 39-52. DOI: 10.1021/ie9003464.

[14] Liu J, Qin T, Yang T, et al. SCR denitration system modeling based on self-adaptive multi-scale kernel partial least squares[J]. *Proceedings of the CSEE*, 2015, **35**: 6083-6088. (in Chinese)

[15] Feng Z, Niu W, Cheng C. Multi-objective quantum-behaved particle swarm optimization for economic environmental hydrothermal energy system scheduling[J]. *Energy*, 2017, **131**: 165-178. DOI: 10.1016/j.energy.2017.05.013.

[16] Farjoudi S, Alizadeh Z. A comparative study of total dissolved solids in water estimation models using Gaussian process regression with different kernel functions[J]. *Environmental Earth Sciences*, 2021, **80**: 1-14. DOI: 10.1007/s12665-021-09798-x.

[17] Xiang L, Wang P, Yang X, et al. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism[J]. *Measurement*, 2021, **175**: 109094. DOI: 10.1016/j.measurement.2021.109094.

[18] Liu Y, Chen H, Wang B. DOA estimation based on CNN for underwater acoustic array[J]. *Applied Acoustics*, 2021, **172**: 107594. DOI: 10.1016/j.apacoust.2020.107594.

基于核极限学习机的 SCR 脱硝系统出口 NO_x 浓度动态建模

马 宁[1] 刘 磊[1] 杨振勇[1] 闫来清[2] 董 泽[3]

([1] 华北电力科学研究院有限责任公司, 北京 100045)

([2] 山西大学电力与建筑学院, 太原 030006)

([3] 华北电力大学河北省发电过程仿真与优化控制技术创新中心, 保定 071003)

摘要:为解决变负荷工况下因模型输入变量较多、相关性大导致模型复杂度增加的问题,提出了一种将核极限学习机(KELM)和主成分分析(PCA)相结合的动态建模方法,并应用于选择性催化还原(SCR)脱硝系统出口处的氮氧化物(NO_x)浓度预测. 首先,将主成分分析应用于输入数据特征信息提取,并将提取信息的当前和过往序列值用作 KELM 模型的输入,以反映 SCR 出口处 NO_x 浓度的动态特征;然后,将 SCR 出口的 NO_x 浓度历史数据作为模型的输入,以提升模型精度;最后,利用优化算法确定模型最优参数. 结果表明,与 GPR、LSTM、CNN 模型相比,所建动态模型的预测误差分别减少约 78.4%、67.6% 和 59.3%,说明该模型结构可靠,能够准确预测 SCR 系统出口 NO_x 浓度.

关键词:选择性催化还原;氮氧化物;主成分分析;核极限学习机;动态模型

中图分类号:TK22