Predicting NOx Emission in Thermal Power Plants Based on Bidirectional Long and Short Term Memory Network

Xiaoqiang Wen(✉) and Kaichuang Li

Department of Automation, Northeast Electric Power University, Jilin 132012, China
148088591@qq.com

Abstract. NOx is one of the main pollutants emitted by thermal power plants. Excessive NOx emissions not only cause many negative impacts on the environment but also cause great harm to human health. Power plant NOx prediction technology has drawn more and more attention from the industry. In this paper, a novel bidirectional long and short term memory network (Bi-LSTM) NOx soft-sensing model is proposed for the first time to dynamically predict NOx emissions from power plants in the form of time series. To get better prediction performance, a univariate model and a multivariate model are constructed for comparative study. Besides, Bi-LSTM and different algorithms are compared in the univariate model. In order to confirm the generalization ability of the model, two sets of NOx values of A and B emission outlets from different power plant historical data is used. The results show that the predictive power of univariate models is better than multivariate models. In univariate models, Bi-LSTM is better than other models. On the two sets of data with sampling intervals of 1 min and 2 min, the mean absolute percentage error (MAPE) could reach 2.105% and 4.45%.

Keywords: NOx Emission · Time Series · Bi-LSTM · Dynamic Prediction

1 Introduction

NOx is one of the main emissions of thermal power plants, and excessive emissions of NOx are very harmful to the environment and human health. NOx emission control is an important challenge in boiler operation [1]. In China, the "Boiler Air Pollutant Emission Standards" issued by the State in 2014 stipulated that the mass concentration of NOx emissions for boilers in key areas shall be no more than 200 mg/m^3 for coal-fired and oil-fired boilers and no more than 150 mg/m^3 for gas-fired boilers. Experts and scholars have developed some technologies to reduce NOx emissions, such as selective catalytic reduction (SCR) and selective non-catalytic reduction (SNCR) [2], less excess air [3], low NOx burners and so on. In order to achieve less NOx emissions, flue gas denitration technology and low NOx combustion technology have been widely adopted by major coal-fired power stations [4].

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 535, pp. 337–357, 2024.
https://doi.org/10.1007/978-3-031-50580-5_30

Currently, SCR denitration is widely used in thermal power plants. The system needs to perform corresponding ammonia injection denitration control by measuring the NOx concentration value. Under appropriate chemical reaction conditions, NOx can be reduced to nitrogen and water by injecting a reducing agent such as ammonia. Excessive injection of ammonia may cause lower NOx emissions but will cause ammonia leakage, secondary pollution to the atmosphere, and the by-product will affect the performance of the equipment. On the contrary, too little ammonia injection may lead to the improper reduction of NOx. Therefore, it is very important to timely and accurately inject the right amount of ammonia into the SCR equipment to achieve low NOx emissions [5].

The NOx concentration in the denitration system was measured using the continues emission monitoring system (CEMS). The measurement of the NOx concentration of flue gas must pass through the heat tracing duct and the analysis cabinet. The flow of the flue gas in the heat tracing duct and the measurement of the concentration in the analysis cabinet require a certain period of time, resulting in a certain delay in the CEMS measurement. The measurement delay will affect the control of the ammonia injection amount of the subsequent denitration system so that the feedforward cannot be responded in time during the ammonia injection amount control process, which enhances the difficulty of ammonia injection amount control. The fluctuation of NOx concentration at the outlet of the denitration system will also bigger. There is an obvious delay between ammonia injection at the SCR outlet and NOX emission, which leads to the SCR denitration system not being able to perform denitration control in a timely and accurate manner [6]. In order to compensate for the effects of CEMS measurement lag and ammonia injection delay, a NOx prediction model can be established to predict NOx emissions at the next moment and achieve timely and accurate denitration.

In recent years, with the rapid development of artificial intelligence algorithms, including artificial neural network (ANN), support vector machine (SVM), extreme learning machine (ELM) and other algorithms. Artificial intelligence has also been more applied to industrial modeling. The models based on these methods have fast calculation speed and high accuracy and are suitable for modeling research of complex systems. For this reason, they are widely used in the prediction and modeling of NOx mass concentration in power plant denitration systems. The ANN was modeled using general regression neural network (GRNN), and the number of input variables was determined using correlation analysis and variance inflation factor (VIF) analysis. The results prove that the selected input variables can get better results than the original input variables [7]. Genetic algorithm (GA) was used to select artificial neural network structure and particle swarm optimization (PSO) for combustion optimization. The results show that the power station significantly reduced NOx emissions (22.5%) [8]. The data collected from the power plant were used to construct a feed-forward back-propagation ANN, which used coal combustion parameters as input and NOx emissions as output. Sequential quadratic programming (SQP) and genetic algorithm (GA) were used as fitness functions to find optimal operating conditions [9]. In addition, other scholars had also used ANN to model NOx emissions [10].

Although neural networks have their unique advantages in NOx predictive modeling, they also have two shortcomings, local optimization and overfitting, which are difficult to completely solve. The support vector machine (SVM) is also a method for solving

nonlinear and complex problems and does not have the above-mentioned defects. Therefore, in recent years, it has been widely used in the establishment and research of NOx emission prediction models for power plant boilers. The least-squares support vector machine supervised learning method was used to model NOx emissions based on the historical real-time data obtained from the 1000 MW DC boiler as input. Through sensitivity analysis, the 33 input variables were reduced to 7 [11]. To the prediction of NOx emission from 660 MW power plant boilers, the selected operating parameters were taken as the input variables, and the particle swarm optimization algorithm was used to fine-tune the hyperparameter of the SVM model. The results showed that the optimized model achieved a good prediction effect [12]. In the absence of mechanism analysis, considering the system delay, a NOx prediction model of forgetting factor online least squares support vector machine was established [13]. In addition, other scholars had also used SVM to predict boiler NOx [14].

In addition to using the most classical ANN and SVM, some scholars had proposed many other excellent algorithms. Three models based on deep belief networks (DBN) were developed and a new data collection method was used to estimate NOx emissions from coal-fired power plants [15]. Based on the traditional online sequential extreme learning machine (OS-ELM), a kind of input data sample increment online sequential extreme learning machine (SIOS-ELM) was proposed [16]. The extreme learning machine (ELM) model was introduced to model the relationship between boiler operating parameters and NOx emission. HS(harmonious search) algorithm was used to optimize the operation parameters. Finally, the goal of NOx reduction was achieved [17]. A dynamic data-driven model of a coal-fired power plant boiler, which is a nonlinear autoregressive neural network with external input (NARX), was designed to estimate both NOx and CO emissions [18]. A Gaussian process (GP) was proposed to reduce NOx emissions from 330 MW boilers. GP was used as a model to simulate the relationship between NOx emission characteristics and boiler operating parameters. The hyperparameters of the model were optimized by genetic algorithms [19]. Using a large amount of real-time operational data collected from the Regulatory Information System (SIS), a new deep learning algorithm called deep bidirectional learning machine (DBLM) was proposed to build a NOx emission model for a 300 MW circulating fluidized bed boiler (CFBB) [20].

Although the above methods show good modeling capabilities, they are not satisfactory in terms of time delay, time sequence, and dynamic characteristics. It is gratifying that long and short term memory network (LSTM) can handle these complex problems well, and has been successfully applied in the fields of natural language processing and image processing [21]. In recent years, LSTM have also been widely used in the field of NOx emissions from power plant boilers. A long and short term memory network was introduced to dynamically predict NOx emission from 660 MW tangential coal fired boiler. The results indicate that the LSTM model has excellent accuracy and generalizability [22]. The models proposed above all require feature selection or transformation of input variables, which is cumbersome and relies on other auxiliary variables. The data processing of the univariate model proposed in this paper is simple, it only needs the

historical data of the target value to complete the prediction, and can achieve a good prediction effect. This article compares the proposed univariate model with the multivariate model and finds that the univariate model has a better predictive effect.

The contribution of this article is as follows, (i) A novel deep learning method Bi-LSTM is used to dynamically predict NOx emissions from power plants in the form of time series for the first time. (ii) The NOx values of two sets of A and B emissions from different power plant data was used to confirm the effectiveness of our proposed model generalization ability. (iii) Some important parameters of models were tuned in an orderly and detailed manner. (iv) The NOx prediction results of the univariate model show that root mean square error (RMSE) and mean absolute percentage error (MAPE) have reached below 24.387 mg/Nm^3 and 4.634%, respectively. The rest of the paper is distributed as follows. Section 2 describes the internal calculation process of basic LSTM cells and the structure of the Bi-LSTM model. In Sect. 3, the overall prediction process of the model and the input strategy of the model is introduced. Section 4 describes the basic situation of the boiler, the data preprocessing and the evaluation criteria of the model. Section 5 develops univariate and multivariate models for two cases. Section 6 presents experimental results and discussion and finally concludes this paper.

2 Model Description

2.1 LSTM Description

A typical architecture of the LSTM layer is presented in Fig. 1. In Fig. 1, each line carries an entire vector, from the output of one node to the input of others. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denotes its content being copied and the copies going to different locations.

LSTM implements temporal memory functions through the opening and closing of gates to prevent gradients from disappearing. Three gate controllers were placed around each LSTM neural network cell, namely, input gate, forget gate and output gate, to form a new computing unit. The input gate controls the input of information, the forgetting gate controls the retention of the cell's historical state information and the output gate control information output. Sigmoid activation function makes the output value of the forgetting gate be [0, 1]. When the output is 0, all the information representing the previous state is discarded by the forget gate. When the output is 1, all the information representing the previous state is retained. The internal operation of LSTM is shown in the following equations:

$$\text{The Input gate}: \ i_t = \delta(W_i x_t + U_i h_{t-1} + V_i c_{t-1} + b_i) \tag{1}$$

$$\text{The forget gate}: \ f_t = \delta(W_f x_t + U_f h_{t-1} + V_f c_{t-1} + b_f) \tag{2}$$

$$\text{Neuron}: \ c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{3}$$

$$\text{The output gate}: \ o_t = \delta(W_o x_t + U_o h_{t-1} + V_o C_t + b_o) \tag{4}$$

$$\text{Output: } h_t = o_t \cdot \tanh(c_t) \tag{5}$$

where c_t is the state of LSTM neural network cells at time t, h_t is the output of the cell at time t. W, U and V are the parameter matrix, b is the bias, and δ is the Sigmoid activation function. Tanh is the hyperbolic tangent activation function. i_t, f_t and ot are input gate, forget gate and output gate at time t, respectively. It can be observed in Eqs. (1)–(5) that the input gate, forget gate and output gate are each connected to a multiplier to control the input and output of information and the state of each cell.

2.2 Bi-LSTM Description

Bi-LSTMs have been widely used in sentiment analysis, classification, biomedicine and speech recognition and have proven to improve the prediction performance of time series. For most forecasting tasks, it is important to be able to know both past and future information within a certain time frame. However, the LSTM hidden layer can only extract some past information, and the Bi-LSTM can train the hidden layer to efficiently extract future and past features at the same time through the forward and backward edges, which can greatly improve the accuracy of the prediction.

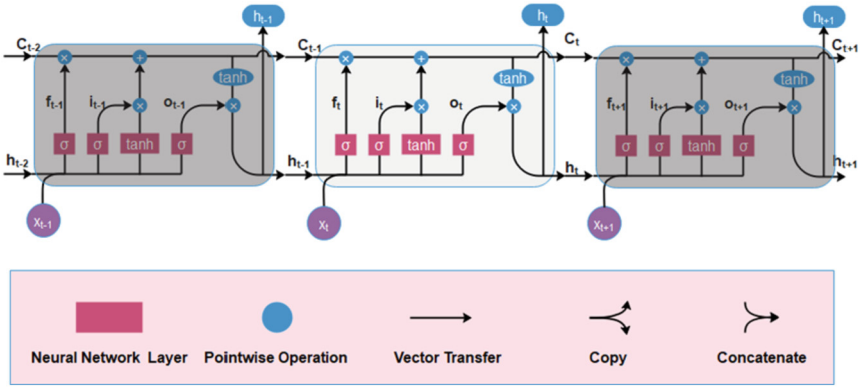


Fig. 1. Internal structure diagram of LSTM cell (Color figure online)

Bi-LSTM predicts the output based on the whole time series. Firstly, the hidden layer neurons are divided into two parts: positive time direction and negative time direction, with two independent hidden layers, and then feedforward to the same output layer, which includes the past and future sequence information at the same time. LSTM of layer 1 calculates the sequence information of the current time point, while LSTM of layer 2 reads the same sequence in reverse and adds the reverse information. The LSTM of each layer has different parameters. The Bi-LSTM structure is shown in Fig. 2.

3 Description Based on Bi-LSTM Prediction Model Framework

3.1 Model Process Description

The prediction process is the same for both univariate and multivariate models. First, raw data were normalized and then divided the data into training and test set. The training data was input into the model, and the parameters in the model were continuously updated through forward propagation and error backpropagation until the Bi-LSTM neural network reaches the number of iterations. Finally, a convergent model was obtained. The test data set was subsequently entered into the convergent model to verify the performance. The detailed flowchart is shown in Fig. 3.

3.2 Prediction Strategies of the Univariate and Multivariate Model

First, the original time series data of the univariate model was converted into input and output variables. The input vector was composed of the previous k NOx emission historical values, and the output vector of the model was the NOx value at the next moment. The measured value at the next moment was used as the input variable for rolling prediction until the whole time series prediction was completed. The prediction strategy of the univariate model is shown in Eqs. (6) and (7).

$$Y_t = \left[y_{t-k}, y_{t-k+1}, \ldots\ldots, y_t\right] \tag{6}$$

$$\hat{y}_{t+1} = f(Y_t) \tag{7}$$

where y_t represents the observed value of the target variable at time t, k represents the time step, Y_t represents the input matrix, f(.) represents the univariate model, and $\hat{y}_{t+1}$ represents the predicted value at time $t + 1$.

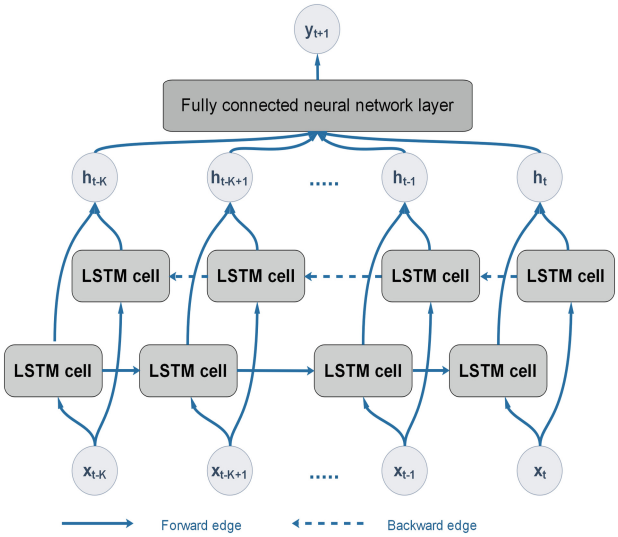


Fig. 2. Structure schematic diagram of Bi-LSTM NOx prediction

The input of the multivariate model was composed of k historical values of the other m variables. Through Pearson correlation analysis, the variables with the absolute value of the correlation coefficient greater than 0.4 were screened. In case 1, 10 historical data of the 21 variables, were used as input variables of the multivariate model. In case 2, 7 historical data of the 11 variables were used as input variables of the multivariate model. The specific extraction auxiliary variables are shown in Table 1 and Table 2. The setting of step size k will be discussed in detail in Sect. 6. The prediction strategy of the multivariate model is shown in Eqs. (8) and (9).

$$X_t = \begin{bmatrix} x_{t-k}^1 & \cdots & x_t^1 \\ \vdots & \ddots & \vdots \\ x_{t-k}^m & \cdots & x_t^m \end{bmatrix} \tag{8}$$

$$\hat{y}_{t+1} = g(X_t) \tag{9}$$

where m and t are the number and time of the auxiliary variable respectively, k represents the time step, x_t^m represents the observation value of the auxiliary variable of the m-th variable at time t, X_t represents the input matrix, g(.) represents the multivariate model, and $\hat{y}_{t+1}$ represents the predicted value at $t + 1$.

4 Data Preparation

4.1 Power Plant Boiler Equipment Description

Two cases were studied in this paper and they are referred to as case 1 and case 2 respectively. Case 1 is a 660 MW square-tangential coal-fired boiler. The boiler is equipped with six medium speed coal mills, which feed pulverized coal into the six sets of burners. Among the six sets of burners, there are six primary air nozzles (A, B, C, D, E, and F) and seven secondary air nozzles (AA, AB, BC, CD, DE, EF, and FF) for fuel combustion air. Eight separated overfire air nozzles (SOFA) and two close-coupled overfire air nozzles (CCOFA, CCOFB) are installed on the left and right walls above the main combustion area. One of the six medium speed coal mills is used as standby.

Case 2 is a 660 MW supercritical pulverized coal boiler. The pulverized system adopts the medium speed direct blowing pulverized coal system, with a total of 7 coal mills, and one as standby. A coal mill corresponds to a layer burner, and seven layers of stationary burner are installed totally. There are seven layers of horizontal thick and light coal powder primary air nozzles (A, B, C, D, E, F, and G), ten layers of secondary air nozzles (AA, OA, OB, CC, DD, DE, EE, FF, FG, and GG) and four layers of separate over fire air nozzles (HH, II, JJ, and KK) in the furnace. For the models established here, all the data are collected from the two boilers above.

4.2 Data Preprocessing

The total amount of data used in each case is 5000. Sample intervals in case 1 and 2 are 1 min and 2 min respectively. The collected data are divided into two parts, namely the

Table 1. Auxiliary variables of multivariate model in case 1

Auxiliary variable	Absolute values of Pearson correlation coefficient
Power	0.75008
Total air volume	0.78532
Primary wind temperature	0.78421
Furnace negative pressure	0.65278
Primary wind pressure	0.7019
Primary air damper A	0.74286
Primary air damper B	0.76168
Primary air damper C	0.72756
Primary air damper D	0.75096
Primary air damper E	0.76279
Primary air damper F	0.6009
Secondary air damper AA	0.70012
Secondary air damper AB	0.71994
Secondary air damper BC	0.71313
Secondary air damper CD	0.67217
Secondary air damper DE	0.67107
Secondary air damper EF	0.63326
Secondary air damper FF	0.82296
Overfire air damper A	0.7338
Overfire air damper B	0.50827
Excess air coefficient	0.64454

Table 2. Auxiliary variables of multivariate model in case 2

Auxiliary variable	Absolute values of Pearson correlation coefficient
Total fuel instruction	0.5068
Total airflow instruction	0.48068
Primary wind pressure	0.40606
B coal feeder speed	0.44723
C coal feeder speed	0.43887
D coal feeder speed	0.41768
F coal feeder speed	0.40027
II over fire air	0.4848
JJ over fire air	0.46319
KK over fire air	0.62019

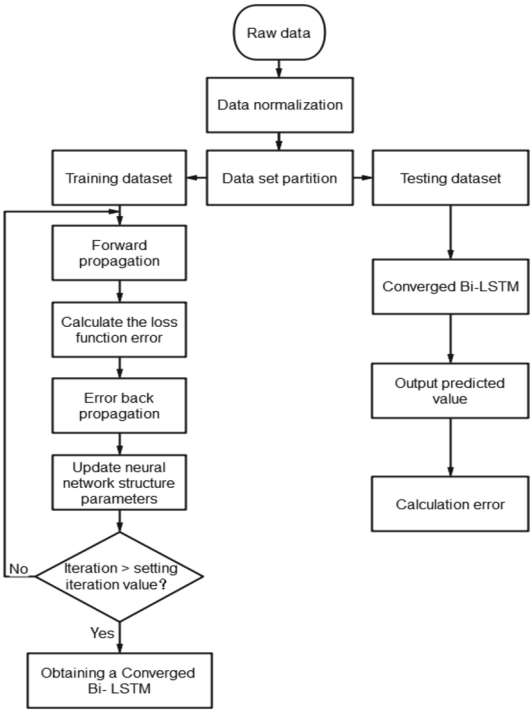


Fig. 3. Flowchart of the univariate and multivariate model

training set and the test set, and the division ratio is 7:3. That is, the training samples are 3500, and the testing samples are 1500. The statistical information of the two sets of data is shown in Table 3.

As can be seen from Table 3, the standard deviation and range in case 2 are both larger than in case 1, indicating that the data in case 2 fluctuates more. It can be seen from the median and average statistics that the NOx on the A and B sides in case 1 has a large difference, and the NOx emissions on both sides of case 2 are more closer.

For the models, the NOx time series was min-max normalized and scaled the value range to the interval [0, 1]. The processed data were used to build the prediction models, and the predicting results were then anti-normalized. The normalization and anti-normalization formulas are presented in Eqs. (10) and (11):

$$x^{(i)} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{10}$$

$$y_i = y^{(i)} * (x_{max} - x_{min}) + x_{min} \tag{11}$$

where xi is the original sequence, xmin is the minimum value of the original sequence, xmax is the maximum value of the original sequence, x(i) is the normalized value, y(i) is the model output value, and yi is the final predicted value.

For a multivariate model, more variables are not always better, and blindly increasing input variables may lead to a decline in the prediction effect of the model. Therefore, in

this paper, Pearson correlation analysis was performed before modeling. Variables with the Pearson correlation coefficients greater than 0.4 were extracted and used as input variables of the model. The calculation formula of the Pearson correlation coefficient is shown in Eq. (12).

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{12}$$

Finally, for case 1 and case 2, 21 and 11 variables were extracted, respectively. All variables are then normalized and anti-normalized.

Table 3. Statistics of the target variables

Data	Sides	Total	Mean	Standard deviation	Minimum	Median	Maximum	Range (Max-Min)
Case 1	A side NOx	5000	385.198	26.698	278.23	386.08	482.63	204.4
	B side NOx	5000	312.504	34.241	217.37	311.33	446.11	228.74
Case 2	A side NOx	5000	370.9	73.316	193.356	367.75	620.75	427.394
	B side NOx	5000	349.635	65.065	197.371	346.444	564.9	367.528

4.3 Model Evaluation Criteria

In this study, three evaluation indicators were used to evaluate the performance of the models, namely root mean square error (RMSE), mean absolute percentage error (MAPE), and R-Square (R2). The definitions of the three formulas are shown in Eqs. (13), (14) and (15):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{13}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - y_i)^2} = 1 - \frac{\mathrm{MSE}(\hat{y}, y)}{\mathrm{Var}(y)} \tag{15}$$

where $\hat{y}_i$ is the predicted value, y_i is the true value, $\bar{y}_i$ is the average of the true values, $\mathrm{MSE}(\hat{y}, y)$ represents the mean square error between the true values and the predicted values, and Var(y) represents the variance of the true values.

5 Model Development

The Bi-LSTM built here is based on the open source artificial neural network library Keras. The loss function of the training model is the mean square error. The time-based backpropagation (BPTT) algorithm, which is similar to the classic backpropagation algorithm, is used as the training algorithm of the LSTM neural network. The Adam algorithm is used as the gradient optimization algorithm, which combines the advantages of adaptive gradient algorithm (AdaGrad) and root mean square backpropagation (RMSProp) algorithm. It can calculate the adaptability of different parameters and occupies fewer processor resources. Compared with other optimization algorithms, Adam algorithm shows great advantages in practical application [44]. Adam is different from traditional stochastic gradient descent (SGD). SGD maintains a single learning rate to update all weights, and the learning rate does not change during the training process. But Adam designs independent adaptive learning rates for different parameters by calculating the first and second moment estimates of the gradient. Therefore, the optimization function adopted in this paper was the Adam algorithm. The initial learning rate of Adam optimization function was set to 0.001. The mean square error is used as the loss function.

In this section, the main hyperparameters of the univariate and multivariate models in case 1 and 2 are tuned. The step size and number of neurons are discussed in detail. For the univariate and multivariate models in case 1 and case 2, a total of 8 combinations are divided. For each combination, the search range for the step size is 1–20 with an interval of 1, and the number of neurons is 20–200 with an interval of 20. In order to prevent the contingency of parameter setting, five repeated tests are carried out for each parameter setting and then the average value is calculated.

5.1 Model Development for Case 1

As can be seen from Fig. 4(a), When the input step is 7, the average RMSE is the smallest. So the step size of the univariate model is set as 7 in case 1. It can be observed in Fig. 4(b) that as the number of neurons increases, the RMSEs also increase, the overall polyline shows an upward trend, and the minimum value appears at the edge. In order to ensure that the best parameters are found, the parameter range is reduced to 5–35 with an interval of 5. Finally, the number of neurons in the univariate model is set as 15 in case 1. It is observed from Fig. 5 that the minimum values in the two figures are 10 and 60 respectively, so the step size and the number of neurons in the multivariate model are set as 10 and 60 respectively.

5.2 Model Development for Case 2

In the same way, the step size and the number of neurons in the univariate model in case 2 are set to 6 and 160, respectively. It can be observed in Fig. 6 and Fig. 7 that the average values of RMSEs fluctuate greatly, and there are several minimum values which are not much different. Finally, it is processed in a simple way. The step size and the number of neurons in the multivariate model are set to 7 and 80, respectively. Finally, the results above are summarized, and the statistical results are shown in Table 4.

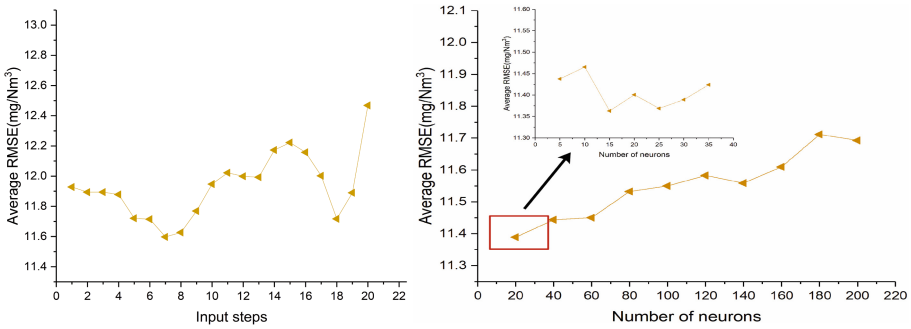


Fig. 4. (a) Tuning the step size of the univariate model in case 1. (b) Tuning the number of internal neurons of the univariate model in case 1.

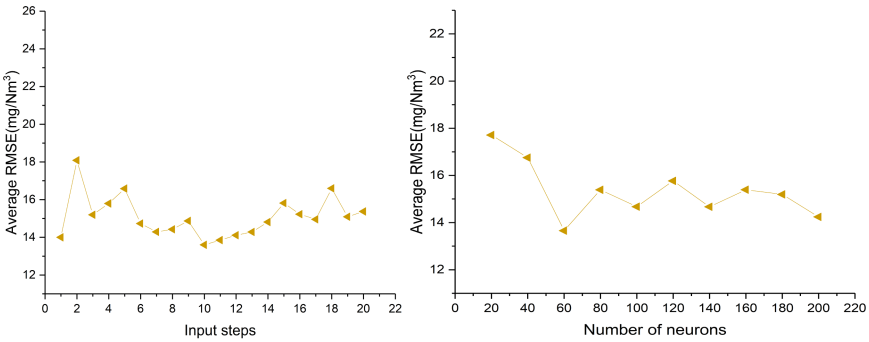


Fig. 5. (a) Tuning the step size of the multivariate model in case 1. (b) Tuning the number of internal neurons of the multivariate model in case 1.

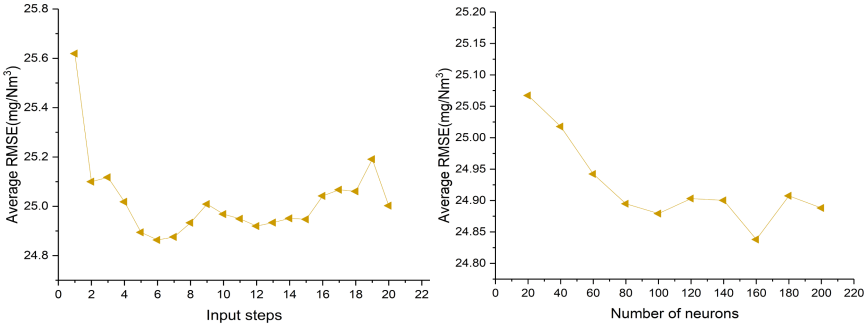
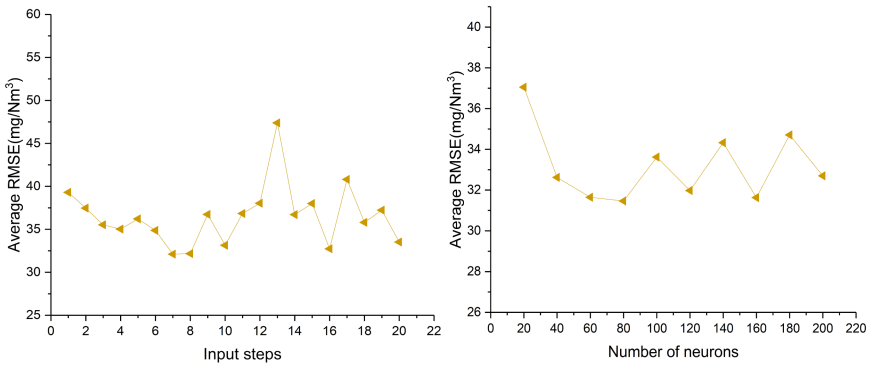


Fig. 6. (a) Tuning the step size of the univariate model in case 2. (b) Tuning the number of internal neurons of the univariate model in case 2.



(a) (b)

Fig. 7. (a) Tuning the step size of the multivariate model in case 2. (b) Tuning the number of internal neurons of the multivariate model in case 2.

Table 4. Tuning of input steps and the number of neurons in univariate and multivariate models in case 1 and case 2

	Case 1		Case 2	
	Input steps	Number of neurons	Input steps	Number of neurons
Univariate model	7	15	6	160
Multivariate model	10	60	7	80

6 Results and Discussion

In this section, The NOx emissions on the A and B sides of the two cases are predicted and compared based on the univariate and multivariate models. The NOx prediction results are presented from three different perspectives and the performance of the models is evaluated by RMSE, MAPE, and R^2.

6.1 NOx Emission Prediction for Case 1

In case 1, the fluctuations of NOx emissions on both sides of A and B are relatively small. Figure 8(a) and (b) are line charts of the predicted and true values corresponding to the univariate models. Figure 8(c) and (d) are the line charts of the predicted and true values corresponding to the multivariate models. The black lines in the four figures represent the true values and the red lines represent the predicted values. It can be clearly seen that the fitting effect of Fig. 8(a) and (b) is obviously better than that of Fig. 8(c) and (d). This intuitively demonstrates that the univariate model is better than the multivariate model. It can be seen from Fig. 8(c) and (d) that the multivariate model can roughly predict the trend of NOx emission, but some details are not well fitted. However, the univariate model can well fit the true values. The four graphs in Fig. 9 present the predicted results from another perspective. The values corresponding to the predicted value and the true

value are represented by red scattered points. The straight line represents the zero error line or perfect line. The closer the scattered points in the figure are to the straight line, the higher the prediction accuracy of the model is. There are some scatter points far from the perfect lines in the four graphs, but the scatter points in Fig. 9(a) and (b) are more concentrated on the straight lines, and the overall trend is the same as the straight line. Figure 9(c) deviates from the perfect line to some extent in terms of the overall trend. Figure 9(d) shows that the scatter points are mainly around the perfect line and more scatter points are far away from the line.

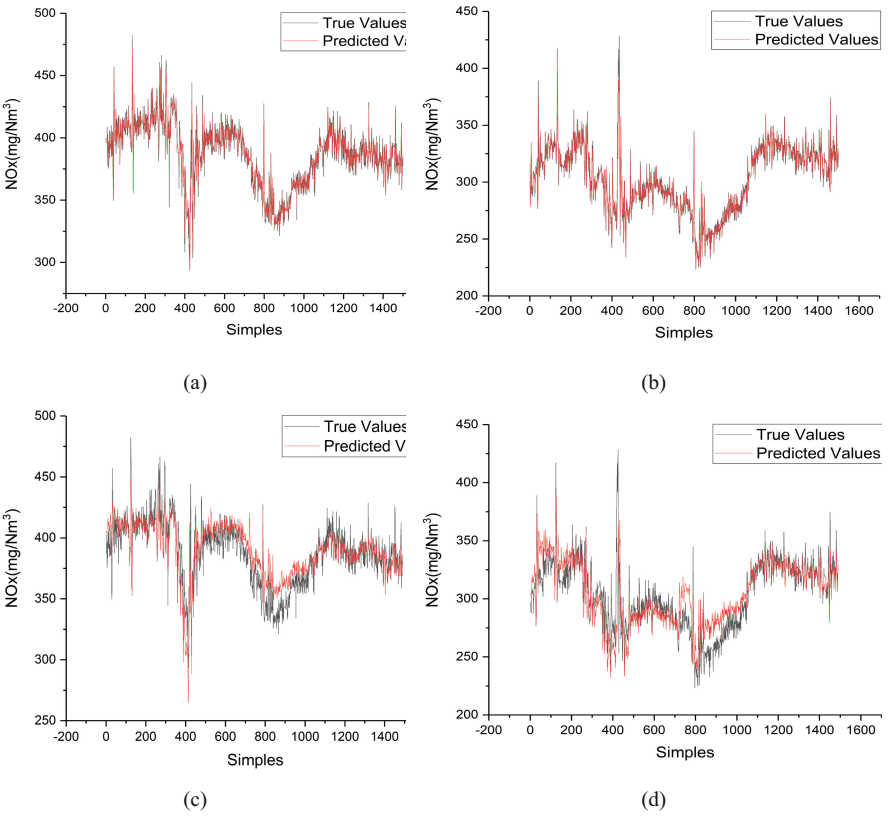


Fig. 8. Line chart of true and predicted values. (a) NOx prediction result on the side A of the univariate model. (b) NOx prediction result on the side B of univariate model. (c) NOx prediction result on the side A of the multivariate model. (d) NOx prediction result on the side B of the multivariate model. (Color figure online)

The residual is the difference between the true value and the predicted value, which can better reflect the prediction effect. In order to show the distribution of the residual more clearly, the number of residuals and the percentage of residuals are counted here. In statistics, the decimal places of the percentage are discarded. As can be seen from Fig. 10(a) and (b), the errors are basically concentrated within ± 10 mg/Nm^3. The closer

the residuals are to zero, the larger the number of statistical samples is. The maximum prediction error is within ± 60 mg/Nm^3. The residuals of Fig. 10(a) and (b) within ± 5 mg/Nm^3 reach 43% and 47% respectively, while the residuals of Fig. 10(c) and (d) are only 30% and 25% respectively. Within ± 10 mg/Nm^3, the former residuals account for 71% and 75% respectively, while the latter for 55% and 49% respectively. The distribution of residuals in Fig. 10(c) is relatively uniform, and the number of samples with residuals near zero is significantly less than that in Fig. 10(a) and (b). The statistical result in Fig. 10(d) is even worse.

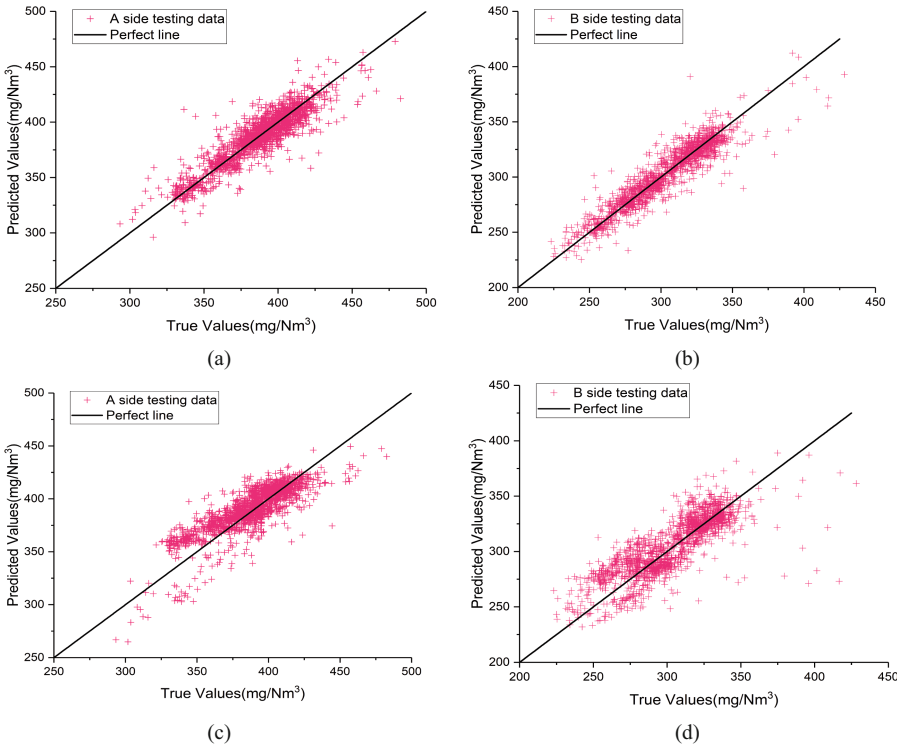


Fig. 9. Scatter plot of true and predicted values. (a) NOx prediction result on the side A of the univariate model. (b) NOx prediction result on the side B of univariate model. (c) NOx prediction result on the side A of the multivariate model. (d) NOx prediction result on the side B of the multivariate model. (Color figure online)

6.2 NOx Emission Prediction for Case 2

In order to verify that the univariate model is better than the multivariate model, we also select a group of data with relatively large fluctuation of NOx emission for prediction. We name it case 2. Similar to case 1, it can be intuitively seen in Fig. 11(a) and (b) that the fitting effect is better. In Fig. 11(c) and (d), the predicted value can better fit

the general trend, but there are some deviations in some time periods. As can be seen from the four graphs in Fig. 12, the scatter plot distribution trend is roughly consistent with the perfect line trend. From the detailed observation, in Fig. 12(a) and (b), the scatter points are densely concentrated on the perfect lines, and in Fig. 12(c) and (d), the scattered points are densely concentrated near the perfect lines. In Fig. 13(a) and (b), the proportions of residuals within ± 10 mg/Nm^3 reach 37% and 42%, respectively, compared to 23% and 35% in Fig. 13(c) and (d). Within ± 20 mg/Nm^3, the residuals of the former account for 66% and 71%, respectively, while those of the latter are 45% and 63% respectively. In Fig. 13(a) and (b), the residual values are concentrated near zero. In Fig. 13(c) and (d), the residual values are more evenly distributed. By comparison, the models prediction effect in Fig. 13(a) and (b) is preferable than that in Fig. 13(c) and (d). From three different perspectives of the prediction results, we can conclude that the prediction results of the univariate models are better than those of the multivariate models.

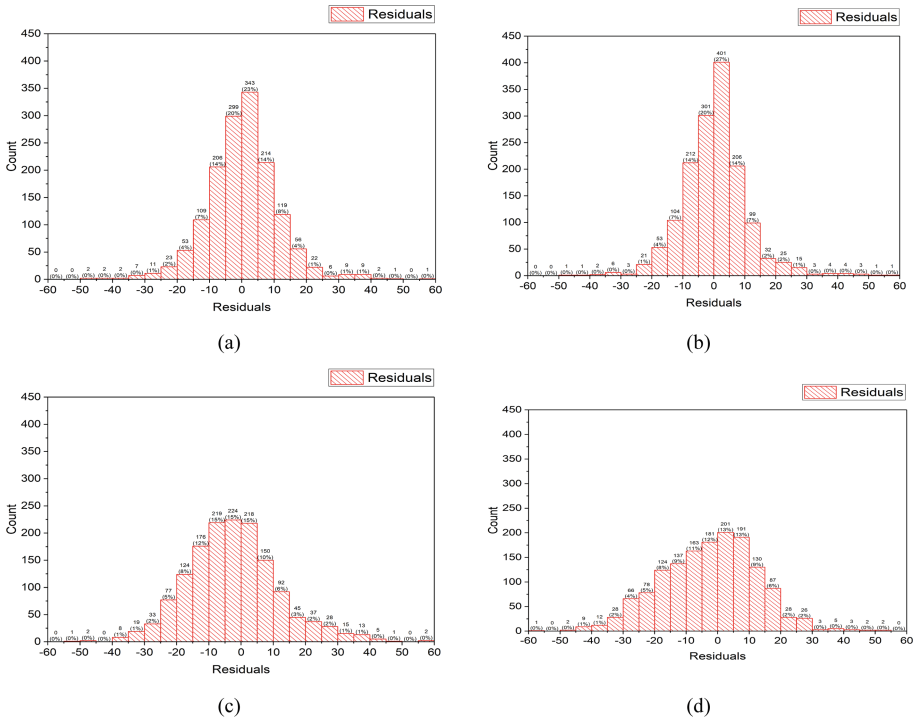


Fig. 10. Residual histograms of true and predicted values. (a) NOx prediction result on the side A of the univariate model. (b) NOx prediction result on the side B of univariate model. (c) NOx prediction result on the side A of the multivariate model. (d) NOx prediction result on the side B of the multivariate model.

We use RMSE, MAPE, and R2 to evaluate the predictive effects of univariate and multivariate models on the A and B sides from the two cases. For the combination of different situations, a total of eight sets of predictions are made, and each set is repeated

five times. Then, the average error of these five times is calculated, and the results are shown in Table 5. As the NOx fluctuation in case 2 is relatively large, the prediction accuracy is not as high as that in case 1. In addition, another reason for the different prediction accuracy of the two sets of data may be that the sampling interval of the two sets of data is different. Case 1 is 1 min, and Case 2 is 2 min. In the same case, the prediction differences on the two sides are not so large, but the prediction errors of different models are very different. In general, for both cases, the univariate model is better than the multivariate model. The average absolute percentage prediction errors of the univariate models can be within 5% in case 1 and case 2. More impressively, the MAPEs of the univariate models are all within 2.5%, and the RMSEs of the A and B side are 11.387 mg/Nm^3 and 10.597 mg/Nm^3 in case 1, respectively, reaching fairly low levels. The R2 values of the four groups of univariate models all reach above 0.8.

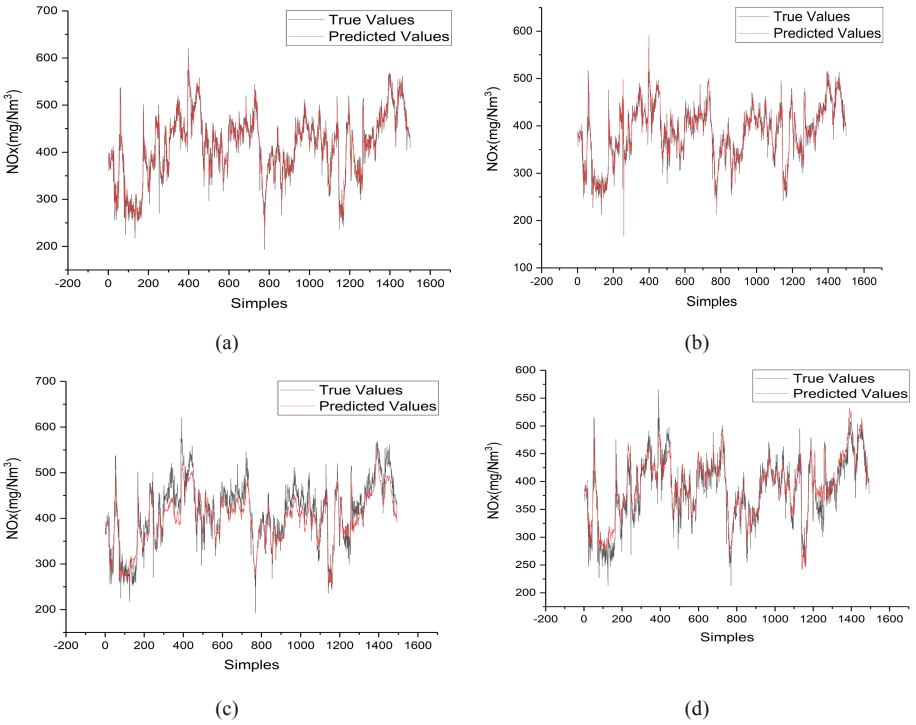


Fig. 11. Line chart of true and predicted values. (a) NOx prediction result on the side A of univariate model (b) NOx prediction result on the side B of univariate model. (c) NOx prediction result on the side A of multivariate model. (d) NOx prediction result on the side B of multivariate model.

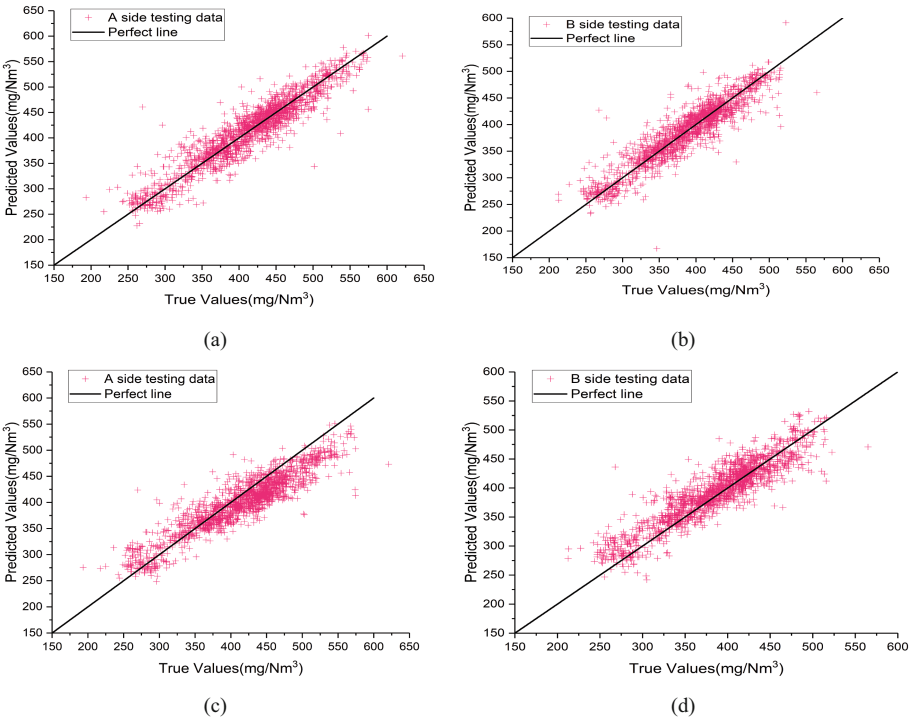


Fig. 12. Line chart of true and predicted values. (a) NOx prediction result on the side A of univariate model. (b) NOx prediction result on the side B of univariate model. (c) NOx prediction result on the side A of multivariate model. (d) NOx prediction result on the side B of multivariate model.

Table 5. Predictive performance of univariate and multivariate models

		Case 1		Case 2	
		Univariate model	Multivariate model	Univariate model	Multivariate model
A side NOx	RMSE (mg/Nm^3)	**11.387**	16.839	**24.867**	31.951
	MAPE (%)	**2.105**	3.423	**4.634**	6.075
	R^2	**0.812**	0.556	**0.87**	0.786
B side NOx	RMSE (mg/Nm^3)	**10.597**	13.926	**23.092**	27.022
	MAPE (%)	**2.469**	3.275	**4.452**	5.451
	R^2	**0.865**	0.764	**0.846**	0.789

6.3 Comparison Between Different Algorithms

This section establishes different comparison models, which are two-layer Bi-LSTM, LSTM, deep neural network(DNN), support vector regression (SVR). The results are shown in Table 6. It can be seen from Table 6 that increasing the number of layers of Bi-LSTM will reduce the prediction performance of the model. The performance of deep learning algorithms LSTM and DNN and machine learning algorithms SVR is general. From Table 6, it is proved that the single-layer Bi-LSTM is superior to other models in different sets of data.

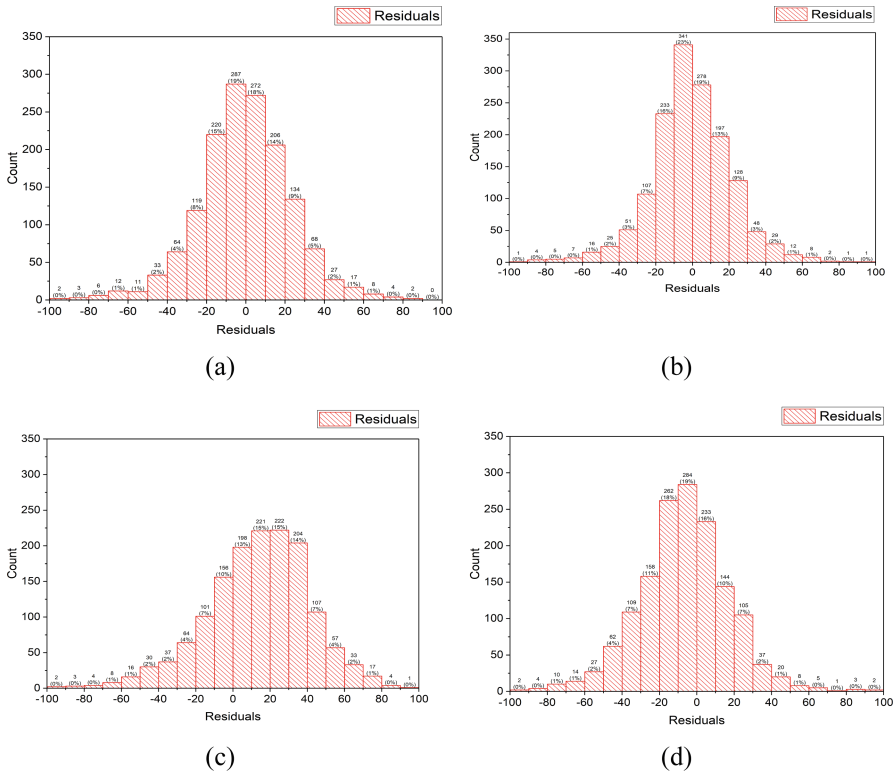


Fig. 13. Residual histograms of true and predicted values. (a) NOx prediction result on the side A of the univariate model. (b) NOx prediction result on the side B of the univariate model. (c) NOx prediction result on the side A of the multivariate model. (d) NOx prediction result on the side B of the multivariate model.

Table 6. Comparison of the results of different models

	Case 1						Case 2					
	RMSE (mg/Nm^3)		MAPE (%)		R^2		RMSE (mg/Nm^3)		MAPE (%)		R^2	
	A side	B side	A side	B side	A side	B side	A side	B side	A side	B side	A side	B side
Bi-LSTM	**11.387**	**10.597**	**2.105**	**2.469**	**0.812**	**0.865**	**24.867**	**23.092**	**4.634**	**4.452**	**0.87**	**0.846**
Bi-LSTM_2	11.773	10.742	2.171	2.470	0.784	0.862	25.003	23.968	4.655	4.541	0.869	0.834
LSTM	12.084	11.949	2.207	2.686	0.773	0.829	27.289	23.789	5.050	4.654	0.843	0.836
DNN	13.744	12.103	2.455	2.704	0.706	0.825	26.089	26.363	4.850	5.060	0.857	0.800
SVR	12.157	11.390	2.256	2.649	0.770	0.845	25.797	23.030	4.761	4.453	0.860	0.847

7 Conclusions

The univariate and multivariate models based on the Bi-LSTM are proposed to predict NOx emissions from two 660 MW coal-fired boilers. The Bi-LSTM structure and prediction process are described in detail and the input strategies of the univariate and multivariate models are introduced, respectively. Through tuning the step size and the number of neurons, the models established achieve the best prediction effect. Finally, the experimental results show that the prediction result of the unilateral model is better than that of the multivariate model. In order to confirm the superiority of the univariate model of the Bi-LSTM algorithm, a comparative study was made here with the univariate models of other algorithms. Experiments confirm that the prediction accuracy of the Bi-LSTM is higher than those of the other algorithms. According to the prediction results, it is found that in the A and B side of case 1, the RMSEs of the Bi-LSTM model is only 11.387 mg/Nm^3 and 10.597 mg/Nm^3, while the MAPEs is only 2.105% and 2.469%, respectively. It can be seen from the above experimental results that this model has good predictive performance and could provide certain guidelines for the SCR denitration system in the thermal power plant. In the future, there will be still much room for improvement in the prediction of time series with large fluctuations by using univariate models. The research in this paper only provides a certain reference for the application of deep learning in thermal power plants.

References

1. Munawer, M.E.: Human health and environmental impacts of coal combustion and post-combustion wastes. J. Sustain. Min. **17**, 87–96 (2018)
2. Zhang, K., Zhao, J., Zhu, Y.: MPC case study on a selective catalytic reduction in a power plant. J. Process. Control. **62**, 1–10 (2018)
3. Nihalani, S.A., Mishra, Y., Juremalani, J.: Emission control technologies for thermal power plants. IOP Conf. Ser. Mater. Sci. Eng. **330**, 012122 (2018)
4. Environmental Protection Agency (EPA): Nitrogen oxides (NOx), why and how they are controlled. EPA-456/F-99-006R. 48 (1999)
5. Xie, P., Gao, M., Zhang, H., Niu, Y., Wang, X.: Dynamic modeling for NOx emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network. Energy **190**, 116482 (2020)

6. Yang, G., Wang, Y., Li, X.: Prediction of the NOx emissions from thermal power plant using long-short term memory neural network. Energy **192**, 116597 (2020)
7. Stamenković, L.J., Antanasijević, D.Z., Ristić, M., Perić-Grujić, A.A., Pocajt, V.V.: Prediction of nitrogen oxides emissions at the national level based on optimized artificial neural network model. Air Qual. Atmos. Heal. **10**, 15–23 (2017)
8. Tuttle, J.F., Vesel, R., Alagarsamy, S., Blackburn, L.D., Powell, K.: Sustainable NOx emission reduction at a coal-fired power station through the use of online neural network modeling and particle swarm optimization. Control. Eng. Pract. **93**, 104167 (2019)
9. Ilamathi, P., Selladurai, V., Balamurugan, K.: A novel approach for modelling of NOx emission reduction in a tangentially fired coal boiler. Int. J. Oil, Gas Coal Technol. **6**, 449–461 (2013)
10. Ilamathi, P., Selladurai, V., Balamurugan, K., Sathyanathan, V.T.: ANN-GA approach for predictive modeling and optimization of NOx emission in a tangentially fired boiler. Clean Technol. Environ. Policy **15**, 125–131 (2013)
11. Wang, Y.L., et al.: Development of a NOx emission model with seven optimized input parameters for a coal-fired boiler. J. Zhejiang Univ. Sci. A **19**, 315–328 (2018)
12. Li, Q., Yao, G.: Improved coal combustion optimization model based on load balance and coal qualities. Energy **132**, 204–212 (2017)
13. Zhai, Y., Ding, X., Jin, X., Zhao, L.: Adaptive LSSVM based iterative prediction method for NOx concentration prediction in coal-fired power plant considering system delay. Appl. Soft Comput. J. **89**, 106070 (2020)
14. Zhou, H., Zhao, J.P., Zheng, L.G., Wang, C.L., Cen, K.F.: Modeling NOx emissions from coal-fired utility boilers using support vector regression with ant colony optimization. Eng. Appl. Artif. Intell. **25**, 147–158 (2012)
15. Safdarnejad, S.M., Tuttle, J.F., Powell, K.M.: Dynamic modeling and optimization of a coal-fired utility boiler to forecast and minimize NOx and CO emissions simultaneously. Comput. Chem. Eng. **124**, 62–79 (2019)
16. Wang, C., Liu, Y., Zheng, S., Jiang, A.: Optimizing combustion of coal fired boilers for reducing NOx emission using Gaussian Process. Energy **153**, 149–158 (2018)
17. Li, G.Q., Qi, X.-B., Chan, K.C.C., Chen, B.: Deep bidirectional learning machine for predicting NOx emissions and boiler efficiency from a coal-fired boiler. Energy Fuels **31**, 11471–11480 (2017)
18. Tai, K.S., Socher, R., Manning, C.D.: Improved semantic representations from tree-structured long short-Term memory networks. In: ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference (2015)
19. Theis, L., Bethge, M.: Generative image modeling using spatial LSTMs. In: Advances in Neural Information Processing Systems (2015)
20. Tan, P., et al.: Dynamic modeling of NOX emission in a 660 MW coal-fired boiler with long short-term memory. Energy **176**, 429–436 (2019)
21. Govoni, N.A.: Generating sequences with RNN. Dict. Mark. Commun. 1–43 (2012)
22. Zhang, X., Kano, M., Matsuzaki, S.: A comparative study of deep and shallow predictive techniques for hot metal temperature prediction in blast furnace ironmaking. Comput. Chem. Eng. **130**, 106575 (2019)