Environmental Science and Pollution Research (2022) 29:13541–13554
https://doi.org/10.1007/s11356-021-16543-1

RESEARCH ARTICLE

Dynamic NO_x emission prediction based on composite models adapt to different operating conditions of coal-fired utility boilers

Guihao Yin [1] · Qinwu Li [2] · Zhongyang Zhao [1] · Lianmin Li [1,3] · Longchao Yao [1] · Weiguo Weng [1] · Chenghang Zheng [1] · Jiangang Lu [4] · Xiang Gao [1]

Received: 24 February 2021 / Accepted: 10 September 2021 / Published online: 30 September 2021

Abstract

An accurate NO_x concentration prediction model plays an important role in low NO_x emission control in power stations. Predicting NO_x in advance is of great significance in satisfying stringent environmental policies. This study aims to accurately predict the NO_x emission concentration at the outlet of boilers on different operating conditions to support the $DeNO_x$ procedure. Through mutual information analysis, suitable features are selected to build models. Long short-term memory (LSTM) models are utilized to predict NO_x concentration at the boiler's outlet from selected input features and exhibit power in fitting multivariable coupling, nonlinear, and large time-delay systems. Moreover, a composite LSTM model composed of models on different operating conditions, like steady-state and transient-state condition, is prosed. Results of one whole day of typical operating data show that the accuracy of the NO_x concentration and fluctuation trend prediction based on this composite model is superior to that using a single LSTM model and other non-time-sequence models. The root mean square error (RMSE) and R^2 of the composite LSTM model are 3.53 mg/m^3 and 0.89, respectively, which are better than those of a single LSTM (i.e., 5.50 mg/m^3 and 0.78, respectively).

Keywords NO_x prediction · CFB boiler · Composite model · Adapt to different conditions · LSTM · Fuzzy c-means

Introduction

Coal accounts for approximately 94% of China's detected fossil energy reserves, and the energy consumption structure dominated by coal is difficult to fundamentally change within a short period. In 2017, China's total energy consumption was 4.49 billion tons of standard coal, with coal consumption accounting for 60.4% of the total energy consumption (China 2018). 83.7%, 63.8%, and 80.1% of the SO_2, NO_x, and particle matter (PM) countrywide, respectively, came from thermal power, steel, and other major energy consuming industries, causing serious air pollution (Liu 2015) and threatening human health. This situation increases the demand for coal fired power plants to decrease pollutant emissions. Circulation fluidized bed (CFB) technology has the advantages of lower NO_x and SO_2 emissions than other widely used combustion processes because of the relatively low bed temperature (ca. 800–900 °C) (Cheng et al. 2020; Tourunen et al. 2009; Koornneef et al. 2007; Basu and Prabir 2006; Weng et al. 2019). In 2019, CFB boilers were responsible for 12% capacity of thermal power plants, which have reached 100,000 MW(Song et al. 2015). Thus, developing the DeNOx technology of CFB boilers has great significance.

Currently, increasingly stringent and detailed NO_x emission limit policies are being set up by governments worldwide for environmental considerations. To meet the rigorous emission standard, most CFB boilers have employed selective catalytic reduction (SCR) or selective non-catalytic reduction (SNCR) technology or both to reduce the NO_x concentration

Responsible Editor: Philippe Garrigues

✉ Chenghang Zheng
zhengch2003@zju.edu.cn

[1] State Key Laboratory of Clean Energy Utilization, State Environmental Protection Center for Coal-Fired Air Pollution Control, Institute for Thermal Power Engineering, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China

[2] Zhejiang HOPE Environmental Protection Engineering Co. Ltd., Hangzhou 310013, China

[3] Jiaxing Xinjia'aisi Thermal Power Co., Ltd., Jiaxing 314000, China

[4] College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

before emission. The key step of these control technologies is determining the amount of reducing agent to use through the NO_x concentration at the spray nozzles. The accurate monitoring of the NO_x concentration in a timely manner currently faces two main difficulties: the variations in boilers' load lead to dramatic NO_x emissions fluctuation and increase the difficulty of NO_x measurement; and the other difficulty results from the delayed measurement of the continuous emission monitoring system (CEMS), which is usually used by power plants for monitoring gas concentrations because of its long heat tracing air extraction pipeline and high-temperature and -dust working environment (Liu 2019).

To address the fast load variation and delayed measurement issues, many researchers have conducted research of developing an accurate method of predicting the NO_x concentration at the boiler outlet. Currently, studies can be mainly divided into two in terms of approach. (1) The first group of studies propose the use of the algorithm based on NO_x production and coal combustion mechanism, which usually predicts NO_x emission using numerical simulation methods (Gungor 2009). G. Lo1ffler et al. (Löffler et al. 2005) studied reducing NO_x emissions from natural gas burners using computational fluid dynamics (CFD) method. In the research of W. J. Sun et al. (Sun et al. 2016), a three-dimensional numerical simulation was carried out to investigate the effect of secondary air, OFA (over-fire air), and AA (additional air) on NO_x emissions and reported that the NO_x concentration is consistent with the measured NO_x concentration. Furthermore, J. Q. Ji et al. (Ji et al. 2020) used a two-dimensional comprehensive CFD combustion model based on the NO_x and N_2O conversion processes and other mechanisms to predict NO_x/N_2O emissions and achieved good prediction accuracy. However, numerical simulation methods are unsuitable in the real-time production environment of a power plant. (2) The second group proposes the use of a data-driven method based on machine learning approaches, such as support vector machine (SVM) and artificial neural networks (ANNs), which is a popular global research direction. The SVM method using nonlinear kernel functions can be strengthened in fitting nonlinear parameters, and some works have been conducted by scholars in Refs.(Ahmed et al. 2015; Tan et al. 2016; Fan et al. 2019; Zhai et al. 2020; Zheng et al. 2008). You Lv et al. (2012) built a nonlinear PLS integrated with LSSVM model to predict NO_x emissions and achieved a root mean square error (RMSE) of 37.6609 mg/m^3 on the test data. ANNs also have excellent capabilities in fitting a nonlinear system (Kalogirou 2003; Zhou et al. 2004; Ilamathi et al. 2013). However, these studies are mainly focused on steady-state conditions while ignoring the transient-state conditions in plant operation. In the transient-state, the model must fit the time sequence data and solve the problem of varying transient-state conditions.

With the development of the data management systems in modern power plants, like the supervisory information system (SIS)(Hong et al. 2020), the information in historical operating data on boilers is increasing and should be utilized for improving operations. One of the big obstacles of utilizing the information is the different time lags between parameters. For example, the fluctuation of coal feeder rate may be much earlier than the change of outlet NO_x concentration. The recurrent neural network (RNN) is a new algorithm that can accommodate consecutive time steps of several parameters (Bengio et al. 1994; Qureshi et al. 2017). To solve the problem of vanishing/exploding gradient in traditional RNN, long short-term memory (LSTM) neural networks, which are capable of dealing with long time-sequential data, were invented (Hochreiter and Schmidhuber 1997). P. Tan et al. (2019) used LSTM neural network to predict the NO_x concentration at a boiler's outlet, and the least RMSE on the test data is 12.2 mg/m^3, which is more accurate than over the 22 mg/m^3 using the SVM method. F. Hong et al. (2020) built a bed temperature sequence interval prediction model for a typical 300-MW CFB unit using the neural network and realized advanced overtemperature warning. P. R. Xie et al. (2020) researched on the NO_x prediction model using a bi-directional LSTM neural network with an attention mechanism, and the mean absolute percentage error was 3.9%. D. Adams et al. (2020) predicted the SO_x and NO_x emissions from a coal-fired CFB boilers with deep neural network (DNN) and LSSVM algorithms and researched on the influence of properties of coal and limestone on SO_x and NO_x emissions, achieving a coefficient of efficiency of 0.8925 and 0.9904 for the predictions of SO_x and NO_x, respectively.

Previous research adopted different methods of predicting NO_x concentration at the boiler's outlet to provide strong support for reducing NO_x emission. However, the operating conditions of boilers change frequently, and the changes are difficult to fit using one universal model. In this manuscript, historical operating data from one CFB unit is studied and divided into different parts according to their operating conditions, like steady-state and transient-state conditions. The LSTM model is separately built to fit the different characteristics of data at different operating conditions, and fast and accurate prediction models compatible with varied changing operating conditions are finally acquired.

Method

CFB boiler structure and data acquisition method

A commercial 220 t/h CFB boiler is selected as the study object. This boiler has a stable coal procurement channel, and the coal quality change factors are ignored in this study. Coal powders ground by in a coal mill and limestone are

mixed together and burned in the boiler. Coal powders with different densities suspend at different heights in the furnace, held by the primary air supplied from the bottom of the bed (Basu 1999). The coal powders that are not burnt are split by the two cyclones at the furnace outlet. Two secondary air nozzles are installed at two sides of the furnace wall, supplying air to complete the combustion. Thermodynamic design parameters of this CFB boiler are listed in Table 1.

Mainly due to the tight space and cost control, the SNCR and SCR combined DeNO$_x$ technology is adopted by this CFB unit. One CEMS is installed at the boiler's outlet after cyclones to monitor the NO$_x$ and O$_2$ concentrations, where the concentration of PM and the temperature of flue gas are both high. In this situation, the CEMS must reversely blow every 2 h to prevent the pumping line from being blocked. During the blowing periods, the measured NO$_x$ concentration drops immediately and presents a huge deviation from the real NO$_x$ concentration, bringing difficulties to NO$_x$ emissions control and requiring auxiliary means of NO$_x$ concentration measurement. Other parameters, such as temperature, flow rate, and pressure, are measured by various meters at different positions, as shown in detail in Figure 1.

All operating data of the thermal plant is stored in the databases of the plant information (PI) system with an interval of 1 s. The raw measured data contains high-frequency noise and takes a long time to reflect the change trend of the boiler. Thus, the moving average method is used to resample the operating data from 1 to 5 s of interval. Furthermore, we fetch two months of operating data from the PI system, that is, approximately 1 million sets of data, as the training data, and 17280 sets data of one day that can represent typical operating conditions as the validation data.

LSTM neural network

The most significant difference between traditional ANN and RNN is that RNN can restore the information of sequential data using the internal state (memory), making RNN more powerful than traditional ANN for fitting sequence inputs. As illustrated in Figure 2 (a), one RNN neuron contains not only the current input but also the previous output. This RNN characteristic indicates the capability of fitting sequence data. Formula (1) shows the computing process:

$$h_t = \sigma\left(\begin{array}{l} W_i x_t + W_c c_{t-1} \\ \qquad\qquad + b \end{array}\right) \tag{1}$$

where h_t represents the output of the hidden RNN neuron at moment t; c_{t-1} is the output of the hidden RNN neuron at moment $t-1$; x_t is the input vector at moment t; W_i is the weight matrix connecting the inputs and the hidden RNN neuron; W_c is the weight matrix connecting two hidden RNN neurons at two near moments; σ is the activation function, which is often the *tanh* function; and b is the bias vector.

However, the traditional RNN faces gradient vanishing or exploding problems. To overcome these problems, the LSTM neural network was proposed by Hochreiter and Schmidhuber (1997), adding three logical memory gates to control the reading, writing, and forgetting operations of each cell. The structure of the LSTM neural network is shown in Figure 2(b) and explained by Formulas (2)-(6) as follows:

$$\text{The input gate}: i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1} + b_i) \tag{2}$$

$$\text{The forget gate}: f_t = \sigma\left(W_f x_t + U_f h_{t-1} + V_f c_{t-1} + b_f\right) \tag{3}$$

$$\text{Neuron output}: c_t = f_t \otimes c_{t-1} + i_t \otimes tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{4}$$

$$\text{The output gate}: o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o c_t + b_o) \tag{5}$$

$$\text{Output}: h_t = o_t \otimes \tanh(c_t) \tag{6}$$

where h_t represents the hidden state of the LSTM cell at moment t; c_t is the cell state of the LSTM cell at moment t; x_t is the input vector at moment t; W_i, W_f, W_c, W_o are weights of the input gate, forget gate, cell output and output gate respectively; σ is the activation function, which is often the tanh(x) function; and b is the bias vector; $\otimes$ represents convolution between two matrixes.

Table 1 Thermodynamic design parameters of the studied CFB boiler

Parameters	Design value
Boiler maximum continuous rating	220 t/h
High superheated steam temperature	540 °C
High superheated steam pressure	13.7 MPa
Low superheated steam temperature	450 °C
Feedwater temperature	242 °C
Feedwater pressure	19.2 MPa
Coal feeder rate of design quality	25.9 t/h

Features selection

To build accurate models, irrelevant features must be abandoned (Safdarnejad et al. 2019). The variables in Figure 1 must be optimized to reduce the network's training time and decrease the dimensions of the input variables, which facilitates the training. Mutual information (MI) is a statistic method that can measure the amount of mutual information between two random features (Yin et al. 2017). The MI factor

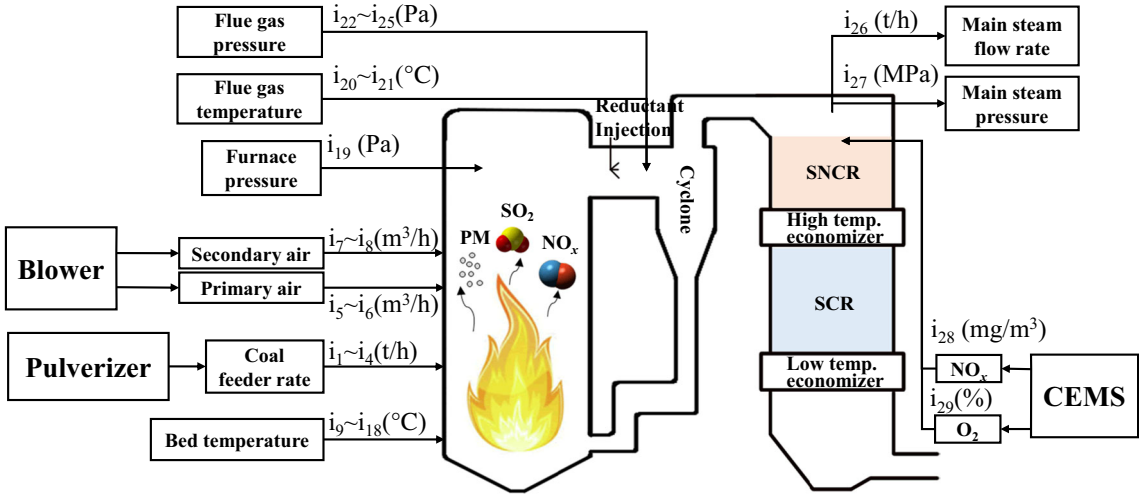

Fig. 1 Schematic and measured parameters of the CFB boiler

ranges from 0 to 1, and the higher the MI factor is, the closer relationship between two features is. Formula (7) defines the mutual information between variables x and y, where $p(x, y)$ is the joint distribution of x and y; $p(x)$ and $p(y)$ are the marginal distributions.

$$I(x,y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x,y)log\frac{p(x,y)}{p(x)p(y)} \tag{7}$$

With several inputs, joint MI is needed and defined as follows:

$$I(x_1,x_2,\ldots,x_n;y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x_1,x_2,\ldots,x_n,y)log\frac{p(x_1,x_2,\ldots,x_n,y)}{p(x_1,x_2,\ldots,x_n)p(y)} \tag{8}$$

where $x_1, x_2, \ldots, x_n$ are the random variables to be selected, and y is the target variable, which is the NO_x emission concentration. F. Wang et al. (2018) used joint MI with the mRMR evaluation criterion (Peng et al. 2005) and the backward elimination algorithm (Guyon and Elisseeff 2003) to calculate the joint MI between the inputs and the NO_x emissions and deleted the moisture in coal. In this study, we calculate the MI between features and the NO_x concentrations, and the results are shown in Table 2.

In Table 2, the furnace pressure, flue gas pressure at furnace outlet, and the main steam pressure have the least MI and are deleted from the input features. The remaining 23 features are selected as the training features.

Delays between features

Along the flue gas flowing path in the boilers, meters are distributed to monitor features, such as the gas temperature, and the gas pressure. However, these features have a time sequence, and the delays between features must be determined before building the models. This paper adopts wavelet transform to find the delays between features.

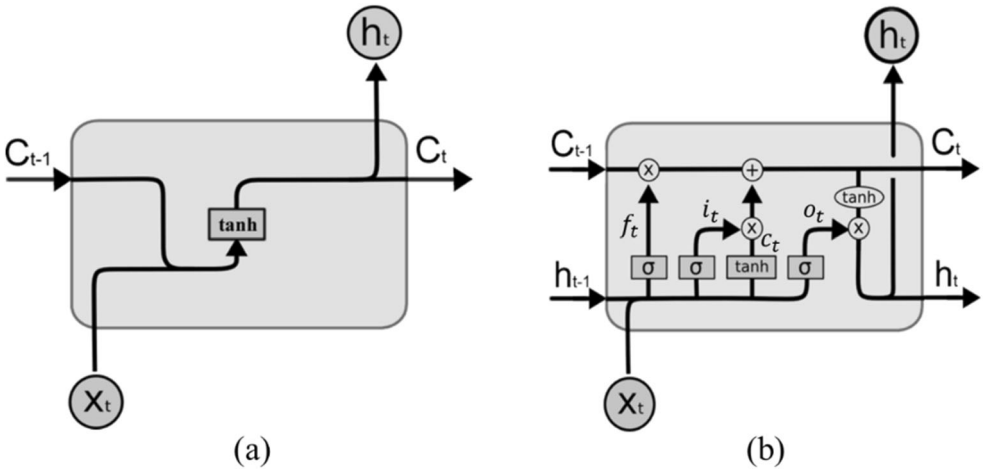

Fig. 2 Structures of one RNN cell (**a**) and LSTM cell (**b**)

Table 2 MI between input variables and NO_x concentration

Parameter name	Identifier	Unit	MI
Coal feeder rate	i_1, i_2, i_3, i_4	t/h	0.53
Primary air flow rate	i_5, i_6	m^3/h	0.36
Secondary air flow rate	i_7, i_8	m^3/h	0.60
Furnace bed temperature	i_9, i_{10}, i_{11}, i_{12}, i_{13}, i_{14}, i_{15}, i_{16}, i_{17}, i_{18}	°C	0.57
Furnace pressure	i_{19}	Pa	0.16
Flue gas temperature	i_{20}, i_{21}	°C	0.68
Flue gas pressure	i_{22}, i_{23}, i_{24}, i_{25}	Pa	0.06
Main steam flow rate	i_{26}	t/h	0.69
Main steam pressure	i_{27}	MPa	0.15
NO_x mass concentration at the boiler's outlet	i_{28}	mg/m^3	1.00
O_2 volume concentration at the boiler's outlet	i_{29}	%	0.80

Wavelet transform is a method of decomposing a signal into a family of wavelet transform functions, which can finally describe the signal (Hao 2010). Through this method, the signal component of the data due to the fluctuation of the instrument itself can be removed, and the delay relationship between signals can be investigated from different scales. As shown in Figure 3 (a), for the raw data from the PI system, the original signal is decomposed into four different frequency band components by wavelet transform, and then the signal deviation value based on the same frequency band is found. Figure 3(b) is the 4th component in the family of wavelets. The minimum point of the three lines marked by "x" is the inflection point of the raw data. Thus, we can determine the delay between the secondary air flow rate, the O_2 concentration, and the NO_x concentration. The rest of the delays between features are determined by the same method, and the results are listed in Table 3.

Table 3 Delays between features determined by wavelet transform

Features	Delay time (s)
Coal feeder rate	120
Primary air flow rate	105
Secondary air flow rate	105
Furnace bed temperature	95
Flue gas temperature	80
Main steam flow rate	70
O_2 volume concentration	70
NO_x mass concentration	0

Model hypermeters

Three hyperparameters, namely, the prediction time step, the look-back time step, and the numbers of nodes of the hidden layers, must be determined to build a suitable LSTM model.

The prediction time step is directly decided by the time lag of the CEMS. The blowing signal of the CEMS of the boiler and the NO_x concentration at the furnace outlet are recorded, and the history data of these two are compared to decide on the prediction time step, as shown in Figure 4. In the picture, the red line representing the CEMS blowing signal ends at the 78th second. If the CEMS is on time, the NOx concentration should immediately increase after the blowing ending. However, the concentration increases at the 150th second. The 72-s gap between these two moments is the time lag of the CEMS and the prediction time step that must be covered. Given 5 s of data sampling interval, the prediction time step is set to 14.

The look-back time step is another LSTM hypermeter and represents the completeness of the historical information (Yang et al. 2020). Finally, the look-back time step is determined as 30.

To adapt the strong nonlinearity characteristic of CFB boilers, two LSTM hidden layers are selected to fit the variables. The numbers of nodes of the two LSTM layers also affect the prediction performance. The particle swarm optimization (PSO) algorithm is utilized here to improve the network structure. The improvement process can be summarized in the following four steps:

(1) Initialize the velocity and location of 40 particles with two dimensions L and N and set the location range randomly from 2 to 30 (nodes number = particle value * 20).
(2) Fit the models using the Adam algorithm with history data and calculate the fitness function.

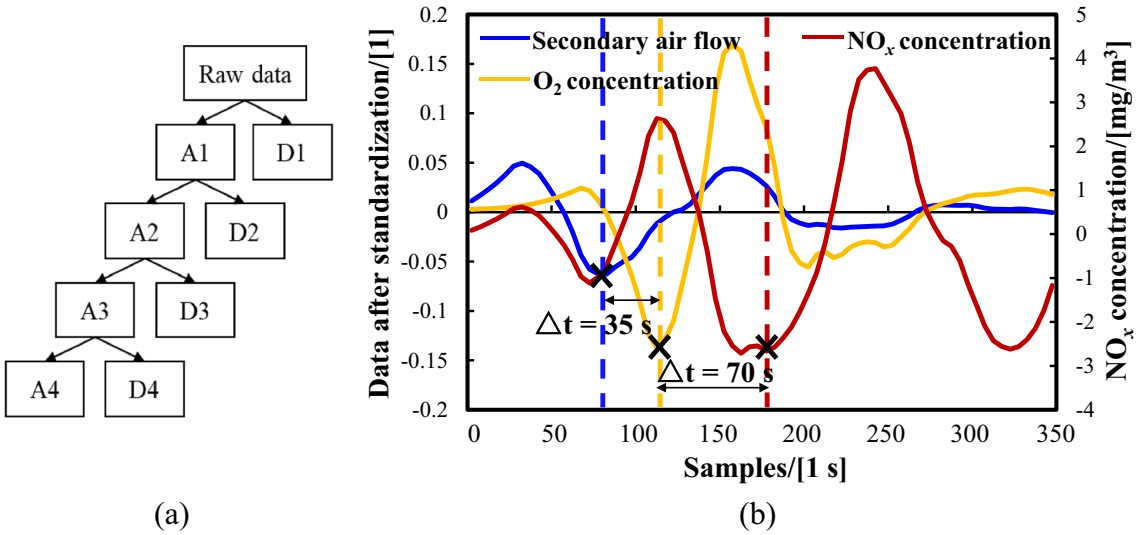

Fig. 3 Determination on delays between features by 4-layer wavelets transformation

(3) Update the velocity and location of each particle using the Formulas (9) and (10), global best position P_i, and local best position P_g. Here, ω is 0.8, and c_1 and c_2 are 2.

$$V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1 \left(P_{id}^{k} - X_{id}^{k}\right) + c_2 r_2 \left(P_{gd}^{k} - X_{id}^{k}\right) \quad (9)$$

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1} \quad (10)$$

(4) Repeat steps (2)–(3) until the number of iterations exceeds 50.

The other hypermeters are set as follows. To avoid overfitting, two dropout layers are separately set after each LSTM layer, and the drop factor is 0.5 according to previous research (Srivastava et al. 2014). The learning rate is set to 0.00166. The optimization process goes through many rounds of iterative and it may take a long time if models are tested on whole dataset. Therefore, data of the first ten days is chosen as the sample in the optimization process. The complete optimization process is depicted in Figure 5. The result of each iteration of PSO improvement is shown in Figure 6, and the best network structure includes two hidden LSTM layers with 220 and 440 nodes. The minimum RMSE on test data is 7.24 mg/m^3, and the result on data of ten days may differ from that on whole sample data.

To nondimensionalize the input features in training the models, standardization is commonly employed in the preprocessing of history data as Formula (11):

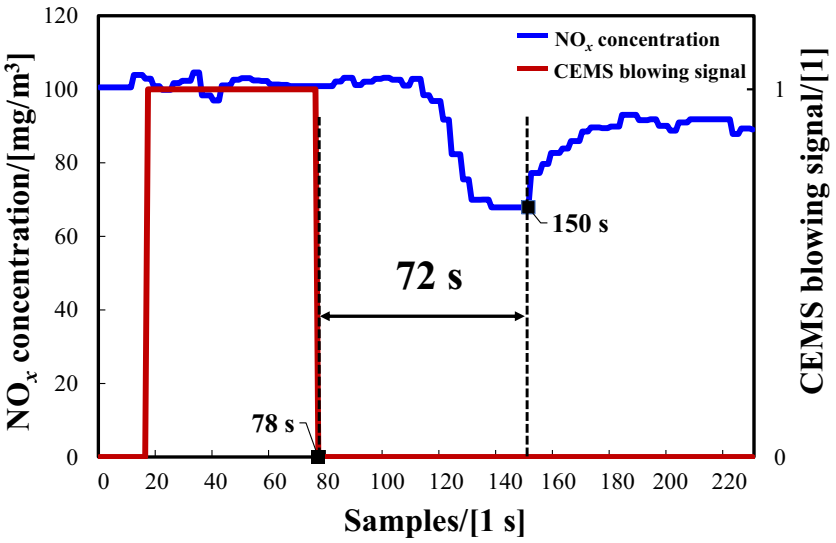

Fig. 4 Determination on prediction time step by analyzing time lag between CEMS blowing signal and NO_x concentration rise

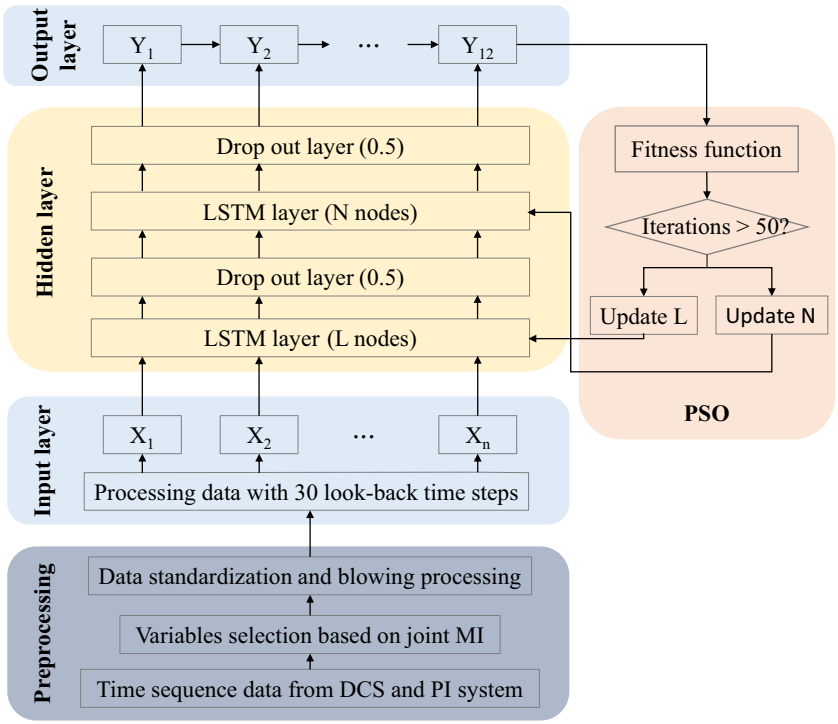

Fig. 5 Optimization of model structure based on the PSO algorithm

$$x^* = \frac{x-\mu}{\sigma} \tag{11}$$

As mentioned, the NO_x concentration is measured by the CEMS at the boiler outlet where the particle concentration is near 200 g/m^3. Thus, the CEMS must blow frequently. To remove the effect of fake measured NO_x concentration during blowing, the history data of NO_x concentration should be preprocessed. First, the variance sliding window moves along the history data of NO_x concentration to mark the start point of blowing, and then linear interpolation is used to process the data during blowing as Formula (12):

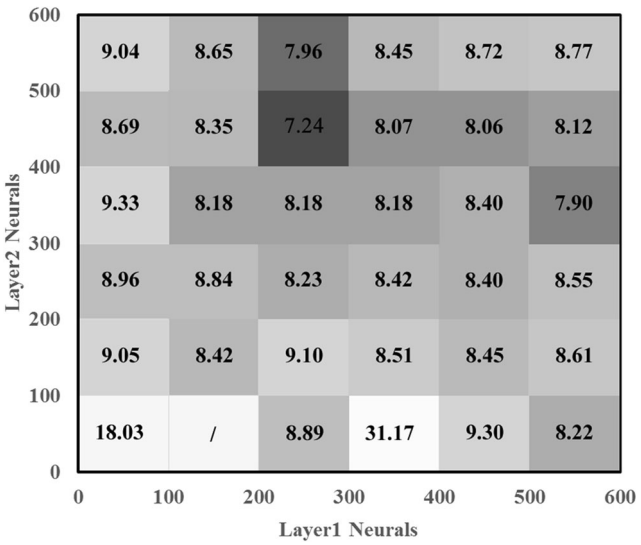

Fig. 6 Result of PSO improvement process

$$\widehat{x_i} = x_{\text{start}} + \frac{x_{\text{end}} - x_{\text{start}}}{(\text{end}-\text{start})} \tag{12}$$

where start represents the start time point of blowing, and end represents the end time point of blowing. x refers to the concentration of NO_x emission.

Model performance index

For the evaluation of the NO_x emissions prediction performance of the models, two indexes, namely, RMSE, and R squared (R^2), are introduced. The two indexes are defined using the following equations:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y_i}-y_i\right)^2}, \tag{13}$$

$$R^2 = 1-\frac{\sum_{i=1}^{n}\left(\widehat{y_i}-y_i\right)^2}{\sum_{i=1}^{n}(y-y_i)^2}, \tag{14}$$

where y_i is the measured value by the CEMS, $\widehat{y_i}$ is the predicted value, and y is the average value of the measured values of the samples.

Table 4 Prediction results of LSTM, DNN, and random forest on two-hour training data and two-hour test data

	LSTM		DNN		Random forest	
	Training data	**Test data**	**Training data**	**Test data**	**Training data**	**Test data**
RMSE	2.54	7.95	11.61	17.11	6.90	17.06
R^2	0.98	0.87	0.61	0.40	0.84	0.40

Results and discussion

DNNs, the random forest, and a LSTM network are compared in terms of their ability to deal with time sequence data and discussed in Section 3.1. In Section 3.2, the composite LSTM neural networks with intelligent data segmentation based on fuzzy c-means clustering is explored to fit the accuracy of prediction when units operate on transient-state condition.

LSTM model compared with other models

For a meaningful comparison, the DNN also has two hidden layers, and each with 220 and 440 nodes, like the LSTM network with the best performance. Random forest is a bagging ensemble algorithm, which has 200 estimators in this study. The prediction results of the LSTM network, the DNN, and random forest are shown in Figure 7 and Table 4.

The RMSEs of the LSTM, the DNN, and the random forest in the training data are 2.54, 11.61, and 6.90 mg/m^3, respectively, and 7.95, 17.11, and 17.06 mg/m^3 in the test data. Therefore, under the same network structure, the one single LSTM neural network exhibits better performance in following and forecasting sequence data than the DNN and random forest. Furthermore, the blue line enclosed by the orange dashed box in Figure 7 is a section of horizontal line, which means that the CEMS blowing period and the measured value are not real. However, the LSTM prediction results still reflect the NO_x concentration value, unaffected by the blowing signal protection, and in some way solve the problem that NO_x cannot be measured during the blowing period.

However, the prediction result of one single LSTM one test data, especially samples 2200–2400 in Figure 7, is not accurate enough. The maximum error of estimated NOx concentration and measured NO_x concentration during this section reaches 25 mg/m^3. It proves that there is still plenty of scope for improvement of the prediction models. On this condition, we study the composite LSTM models and the models are discussed in the next section.

Composite LSTM model on different operating conditions

The previous section proves that the LSTM network is stronger at coping with sequence data than the DNN and the random forest models. However, when the unit load fluctuates frequently, the network sometimes reacts slowly. Given this difficulty, we studied typically historical operating data segmentation based on the fuzzy c-means clustering algorithm and build composite LSTM model including LSTM model fitting well on certain operating condition to enable the global prediction model better compatible with various operating conditions.

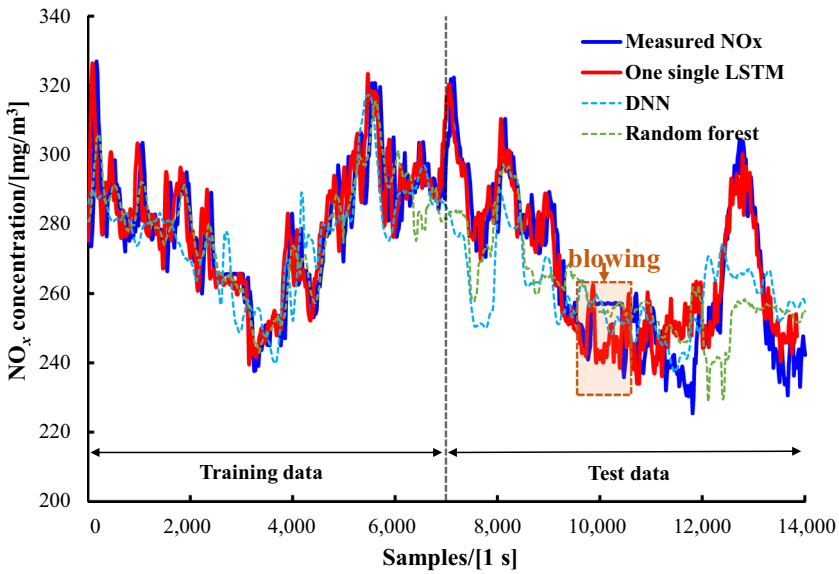

Fig. 7 Prediction results of one single LSTM, DNN, and random forests

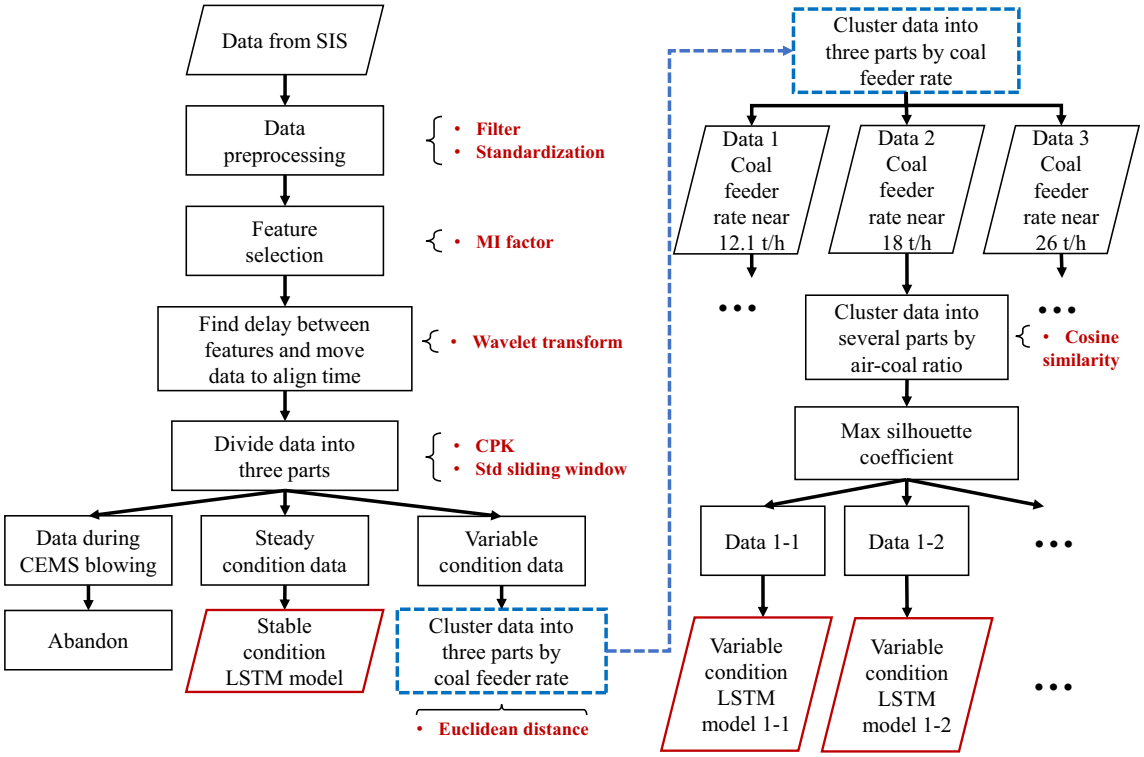

Fig. 8 Main steps of building composite model for predicting NO_x concentration at boiler's outlet

The operating conditions of a boiler can be divided into two major kinds, namely, steady-state and transient-state operating conditions. The flow chart of building composite LSTM model and related methods are shown in Figure 8.

CEMS blowing condition

During CEMS blowing, the NO_x concentration measured by the CEMS is deviated from true value. To avoid the fault information been stored in database, the plant operators compulsively turn fluctuated curve of boiler's outlet NO_x concentration into a straight line during CEMS blowing. That means the standard deviation (SD) of measured NOx concentration during CEMS blowing is much lower than normal operating conditions. Therefore, a SD sliding window is utilized to monitor whether the sample data of NO_x concentration is lower than usual to identify condition during CEMS blowing. Because NO_x concentration data during CEMS blowing is fictitious measured value, operating data during CEMS blowing is abandoned from modeling.

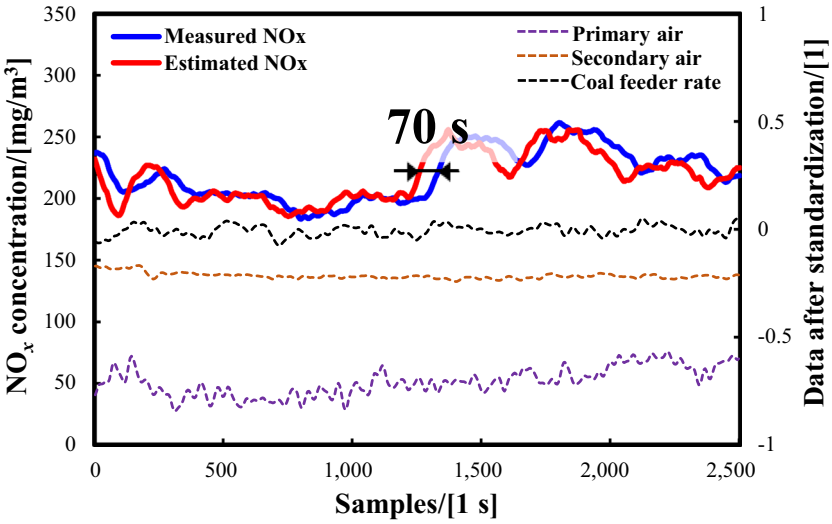

Fig. 9 Prediction result of NO_x concentration by the LSTM model in steady-state operating condition on test data

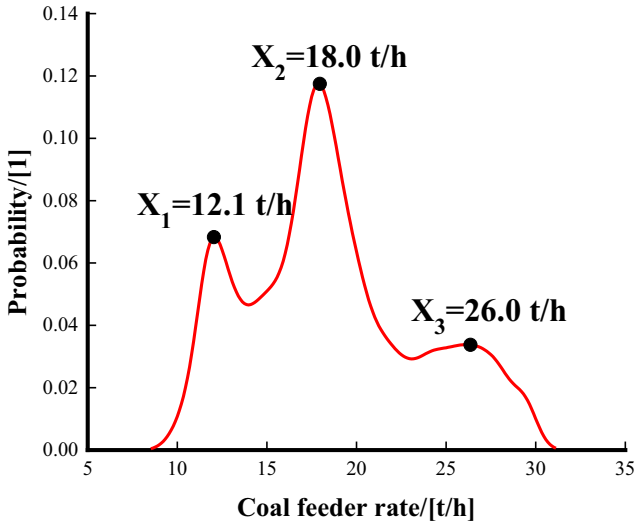

Fig. 10 Distribution of total coal feeder rate based on kernel density estimation

Steady-state condition

First, steady-state operating conditions are intercepted from two months of data based on the CPK method. This method is used in the industry to represent the ability of keeping process stable measured by the mean and standard derivation of data. Here, we define that coal feeder rate, total air, and main steam pressure of the boiler keeps continually between [mean − 1.5 * Std., mean + 1.5 * Std.] at least 2 h as one section of steady-state operating condition. The steady-state condition LSTM model is trained using steady-state operating data, and the prediction result on the test data is shown in Figure 9. The RMSE is 4.86 mg/m^3, and R^2 is 0.82.

Transient-state condition

The rest of the data, apart from the data in the blowing period, is all under the dynamic operating condition. The coal feeder rate and the total wind amount are commonly known as fundamental primary factors that represent the conditions of boilers and the influence NO_x emissions. The coal feeder rate determines the thermal load of boilers, which finally determines the unit and heat loads of boilers. Furthermore, the coal feeder rate and total air amount versus total coal feeder rate, defined as air-coal ratio, together determine the NO_x variation. For example, when the unit must increase the load, operators usually increase the primary and secondary air flow rate first before the coal feeder rate, which increases the air-coal ratio. Consequently, the NO_x concentration also increases immediately. Thus, we select the coal feeder rate and air-coal ratio as the criterion of data clustering. The kernel density estimation and the fuzzy c-means clustering algorithm are adopted, and the main procedure is presented as follows:

1) The outer layer divides the data into several sections based on the kernel density estimation, which is used to determine the typical operation points of the coal feeder rate from several loads with the maximum probability. P. Emanuel et al. (Parzen 1962) provided the detailed method and result of kernel density estimation using the Gaussian kernel function, as shown in Figure 10. When the coal feeder rate equals 12.1, 18.0, and 26.0 t/h, the density reaches the regional maximum and thus divides the training data into three sections.

2) The inner layer works further on each data group divided by the outer layer and then divides each group into different parts based on the cosine distance of the air-coal ratio of every data batch using the fuzzy c-means clustering method. Fuzzy c-means is an unsupervised clustering method, which can automatically find the cluster center on a certain number of clusters. Here, the cluster index is cosine similarity, and the cluster accuracy valuation index is silhouette coefficient (S_c), which are expressed as follows:

$$S(i) = \frac{b(i) - a(i)}{\max(a, b)}, \tag{15}$$

$$S_{\cos}(x_i, x_j) = \frac{\sum_{q=1}^{n} (x_{iq} \times x_{jq})}{\sqrt{\sum_{q=1}^{n} (x_{iq})^2} \times \sqrt{\sum_{q=1}^{n} (x_{jq})^2}}, \tag{16}$$

where $S(i)$ is the silhouette coefficient of one sample, $a(i)$ is the similarity of one sample with the other samples in the same cluster, and $b(i)$ is the similarity of one sample with the other samples in different clusters. The range of $S(i)$ is [-1, 1], and the nearer $S(i)$ is to 1, the better the cluster performance will be. $S_{\cos}(x_i, x_j)$ is the cosine similarity of two samples.

The cluster results are shown in Figure 11, and the average of silhouette coefficient on different numbers of clusters is provided in Table 5. In Figure 11, the horizontal axis marks three different coal feeder rate conditions, and the difference

Table 5 The silhouette coefficient of each cluster

Silhouette coefficient	Coal feeder rate		
	12.1 t/h	18 t/h	26 t/h
Air coal ratio			
2 clusters	0.722	0.767	0.744
3 clusters	0.811	0.812	0.736
4 clusters	0.705	0.685	0.423
5 clusters	0.66	0.689	0.385

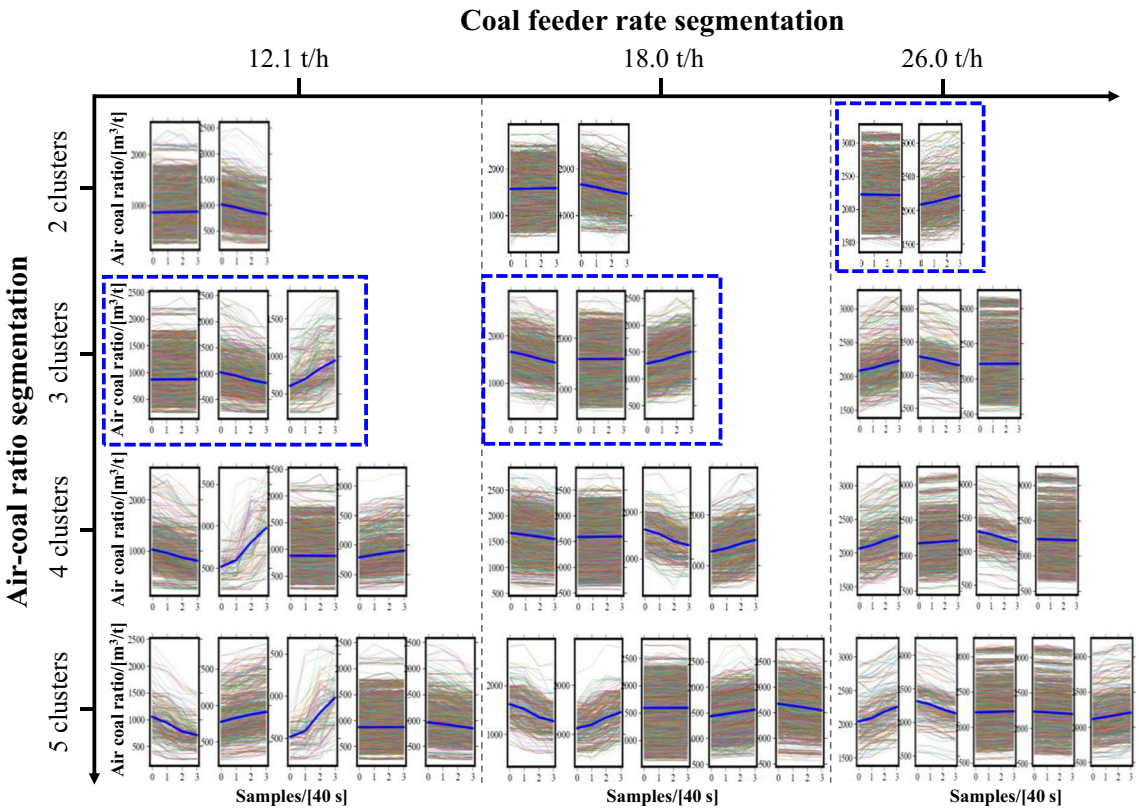

Fig. 11 Clustering results based on two layers fuzzy c-means clustering method. Blue boxes represent the clustering numbers

in fuel means different boiler operating loads. Thus, the bed temperature and the base line of the NO_x concentration are different. Then, the vertical axis marks different numbers of clusters by fuzzy c-means. The bolder blue line in each picture represents the cluster center line, which has the maximum cosine similarity with the lines in each picture. Moreover, in every condition of clustering, the gradient of the air-coal ratio line is the clustering index, such as increasing sharply, increasing slowly, stable, decreasing slowly, and decreasing sharply. Each picture represents one kind of boiler operating condition.

Table 6 Performance of transient-state condition LSTM models for NO_x prediction validated on one-hour validation data

RMSE (mg/m^3)	Coal feeder rate clusters		
	12.1 t/h	18 t/h	26 t/h
Air-coal ratio clusters			
Cluster I	2.32	1.25	2.09
Cluster II	2.13	1.82	2.01
Cluster III	2.10	2.34	/

After data cluster preprocessing, the data in different parts are trained in different LSTM networks with the same structure. One-hour validation data of transient-state conditions are used to validate the prediction performance of LSTM network models with cluster preprocessing. The RMSE of each model is shown in Table 6. When the main steam flow rate changes sharply, the prediction performance of the NO_x concentration is proven good. The average RMSE and R^2 in the training data are 1.60 mg/m^3 and 0.99, and 3.53 mg/m^3 and 0.89 in test data, respectively.

Tested on all operating conditions

Finally, the steady-state and transient-state condition models are tested in one whole day of continuous validation data covering typical operating conditions, to test the prediction performance of the composite LSTM model. The model selection process at a certain time is the same as the data segmentation and model training process shown in Figure 8. That is, one section of data must be judged by the CPK method to determine whether the operating conditions are steady-state or transient-state. If the conditions are on steady-state, the Euclidian distance of the coal feeder rate from 12.1, 18, and 26 t/h must be calculated and then compared with the cosine

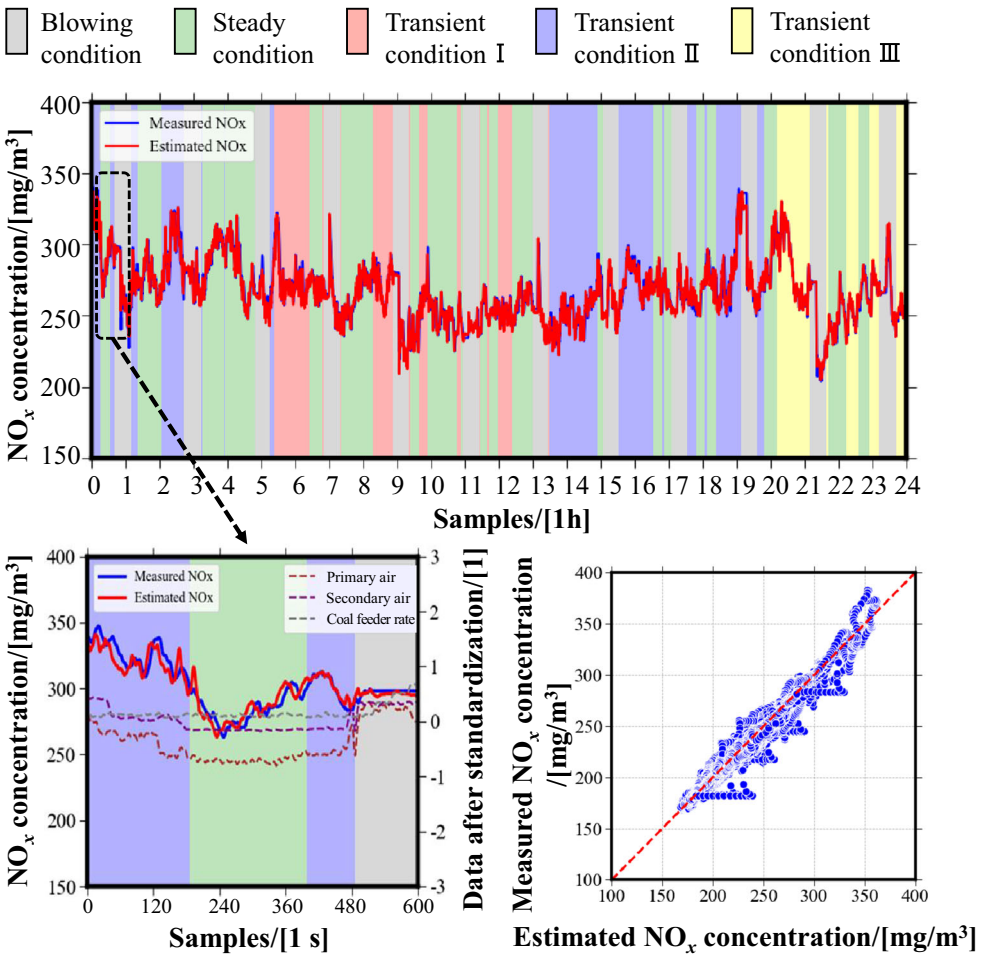

Fig. 12 NO_x prediction with composite LSTM model on different types of operating conditions

similarity of each cluster center. Afterward, the most suitable model is selected from the LSTM models and used to predict the NO_x concentration. As depicted in Figure 12 (top), the different background color at each time represents the operating condition identified by the composite models. Figure 12(bottom) shows the detailed prediction performance, and the model can predict in advance and with high accuracy, proving the ability of built composite LSTM models to fit well in a long period and in different operating conditions. Table 7 shows the results of different predictive models researched in this manuscript, and it proves the composite LSTM model have better ability to fit different operating conditions.

Conclusion

In this manuscript, the structure of hidden layers of LSTM model is optimized by PSO algorithm. Then, the optimized LSTM model is built to compare with DNN and random forests models, proving that the LSTM network is better in fitting sequence data than the other two models. However, the accuracy is not good enough when the load changes in a wide range and cannot quickly reach the load regulation demand of modern power plants. Thus, the composite LSTM model with data clustering preprocessing is researched. The two-layer cluster method divides data into several parts according to the coal feeder and

Table 7 RMSE and R^2 of prediction of different models

	Composite LSTM		LSTM		DNN		Random forest	
	Training data	**Test data**	**Training data**	**Test data**	**Training data**	**Test data**	**Training data**	**Test data**
RMSE	1.60	3.53	2.54	7.95	11.61	17.11	6.90	17.06
R^2	0.99	0.89	0.98	0.87	0.61	0.40	0.84	0.40

air-coal ratio change rates. The data after segmentation is used to train the composite LSTM models, and it exhibits better performance than LSTM without data preprocessing, as well as the DNN and the random forest models.

Author contribution Guihao Yin was responsible for designing experiments and writing the manuscript. Qinwu Li built the database system to fetch data from PI system. Lianmin Li analyzed and preprocessed the data. Zhongyang Zhao and Guihao Yin wrote programs. Longchao Yao debugged the programs. Weiguo Weng and Jiangang Lu analyzed results and helped draw conclusions. Chenghang Zheng and Xiang Gao gave instructions on improving the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National Natural Science Foundation of China (U1609212), Development Plan of Shandong Province of China (2019JZZY010403).

Data availability The datasets analyzed during the current study are not publicly available for the sake of safety and privacy of power plant related, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Adams, Derrick, Dong-Hoon Oh, Dong-Won Kim, Chang-Ha Lee, and Min Oh. 2020. 'Prediction of SOx–NOx emission from a coal-fired CFB power plant with machine learning: plant data learned by deep neural network and least square support vector machine', *Journal of Cleaner Production*, 270.

Ahmed F, Cho HJ, Kim JK, Seong NU, Yeo YK (2015) A real-time model based on least squares support vector machines and output bias update for the prediction of NOx emission from coal-fired power plant. *Korean Journal of Chemical Engineering* 32:1029–1036

Basu, and Prabir. 2006. 'Combustion and gasification in fluidized beds || Stoichiometric Calculations', 10.1201/9781420005158: 445-51.

Basu P (1999) Combustion of coal in circulating fluidized-bed boilers: a review. *Chemical Engineering Science* 54:5547–5557

Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. *Ieee Transactions on Neural Networks* 5:157–166

Cheng L, Ji J, Wei Y, Wang Q, Fang M, Luo Z, Ni M, Cen K (2020) A note on large-size supercritical CFB technology development. *Powder Technology* 363:398–407

National Bureau of statistics of the people's Republic of China. 2018. "Statistical bulletin of the People's Republic of China on national economic and social development in 2017." In, edited by National Bureau of statistics of the people's Republic of China. National Bureau of statistics of the people's Republic of China.

Gungor A (2009) Simulation of NOx emission in circulating fluidized beds burning low-grade fuels. *Energy & Fuels* 23:2475–2481

Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *Journal of Machine Learning Research* 3:1157–1182

Hao ZL (2010) Research on recognition and diagnosis of utility boiler combustion state. *North China electric power university dissertation*

Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Computation* 9:1735–1780

Hong, Feng, Dongteng Long, Jiyu Chen, and Mingming Gao. 2020. 'Modeling for the bed temperature 2D-interval prediction of CFB boilers based on long-short term memory network', *Energy*, 194.

Ilamathi P, Selladurai V, Balamurugan K, Sathyanathan VT (2013)ANN-GA approach for predictive modeling and optimization of NOx emission in a tangentially fired boiler. *Clean Technologies and Environmental Policy* 15:125–131

Ji J, Cheng L, Wei Y, Wang J, Gao X, Fang M, Wang Q (2020) Predictions of NOx/N2O emissions from an ultra-supercritical CFB boiler using a 2-D comprehensive CFD combustion model. *Particuology* 49:77–87

Koornneef J, Junginger M, Faaij A (2007) Development of fluidized bed combustion—an overview of trends, performance and cost. *Progress in Energy and Combustion Science* 33:19–55

Liu, B. W. 2019.

Liu, J. Q. 2015. "Annual report on environment development of China." In.

Löffler G, Sieber R, Harasek M, Hofbauer H, Hauss R, Landauf J (2005) NOxFormation in natural gas combustionevaluation of simplified reaction schemes for CFD calculations. *Industrial & Engineering Chemistry Research* 44:6622–6633

Lv Y, Liu J, Yang T (2012) Nonlinear PLS integrated with error-based LSSVM and its application to NOx modeling. *Industrial & Engineering Chemistry Research* 51:16092–16100

Parzen E (1962) On estimation of probability density function and mode. Annals of Mathematical Statistics. *Ann. Math. Statis.* 33:1065–1076

Peng HC, Long FH, Ding C (2005) Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 27:1226–1238

Qureshi AS, Khan A, Zameer A, Usman A (2017) Wind power prediction using deep neural network based meta regression and transfer learning. *Applied Soft Computing* 58:742–755

Kalogirou SA (2003) Artificial intelligence for the modeling and control of combustion processes: a review. *Prog Energ Combust* 6:515–566

Safdarnejad SM, Tuttle JF, Powell KM (2019) Dynamic modeling and optimization of a coal-fired utility boiler to forecast and minimize NOx and CO emissions simultaneously. *Computers & Chemical Engineering* 124:62–79

Song C, Li M, Zhang F, He Y-L, Tao W-Q(2015) A data envelopment analysis for energy efficiency of coal-fired power units in China. *Energy Conversion and Management* 102:121–130

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15:1929–1958

Sun W, Zhong W, Yu A, Liu L, Qian Y (2016) Numerical investigation on the flow, combustion, and NOX emission characteristics in a 660 MWe tangential firing ultra-supercritical boiler. *Advances in Mechanical Engineering* 8:168781401663072

Tan P, Xia J, Zhang C, Fang Q, Chen G. 2016. 'Modeling and reduction of NOx emissions for a 700 MW coal-fired boiler with the advanced machine learning method', *Energy*: 672-9.

Tan P, He B, Zhang C, Rao DB, Li SN, Fang QY, Chen G (2019) Dynamic modeling of NOx emission in a 660 MW coal-fired boiler with long short-term memory. *Energy* 176:429–436

Tourunen A, Saastamoinen J, Nevalainen H (2009) Experimental trends of NO in circulating fluidized bed combustion. *Fuel* 88:1333–1341

Wang F, Ma S, He W, Li Y, Zhang J (2018) Prediction of NOX emission for coal-fired boilers based on deep belief network. *Control Engineering Practice* 80:26–35

Fan W, Si F, Ren S, Yu C, Cui Y, Wang P (2019) Integration of continuous restricted Boltzmann machine and SVR in NOx emissions prediction of a tangential firing boiler. *Chemometrics and Intelligent Laboratory Systems* 195:103870

Weng WG, Liu BW, Guo YS et al (2019) Model predictive control utilizing gas soft-sensors on SCR denitration system in thermoelectric unit. *BOILER MANUFACTURING* 5:17–22

Xie, Peiran, Mingming Gao, Hongfu Zhang, Yuguang Niu, and Xiaowen Wang. 2020. 'Dynamic modeling for NOx emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network', *Energy*, 190.

Yang G, Wang Y, Li X (2020) Prediction of the NO emissions from thermal power plant using long-short term memory neural network. *Energy* 192:116597

Yin ZL, Li J, Zhang Y, Ren AF, Von Meneen KM, Huang LY (2017) Functional brain network analysis of schizophrenic patients with positive and negative syndrome based on mutual information of EEG time series. *Biomedical Signal Processing and Control* 31: 331–338

Zhai, Yongjie, Xuda Ding, Xiuzhang Jin, and Lihui Zhao. 2020. 'Adaptive LSSVM based iterative prediction method for NOx concentration prediction in coal-fired power plant considering system delay', *Applied Soft Computing*, 89.

Zheng L, Zhou H, Wang C, Cen K (2008) Combining support vector regression and ant colony optimization to reduce NOx emissions in coal-fired utility boilers. *Energy & Fuels* 22:1034–1040

Zhou H, Cen KF, Fan JR (2004) Modeling and optimization of the NOx emission characteristics of a tangentially fired boiler with artificial neural networks. *Energy* 29:167–183