

An Ensemble Deep Belief Network Model Based on Random Subspace for NO_x Concentration Prediction

Yingnan Wang,* Guotian Yang, Ruibiao Xie, Han Liu, Kai Liu, and Xinli Li

Cite This: *ACS Omega* 2021, 6, 7655–7668

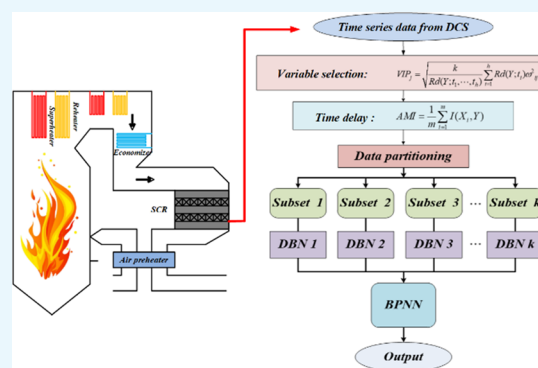
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: An effective NO_x prediction model is the basis for reducing pollutant emissions. In this paper, a real-time NO_x prediction model based on an ensemble deep belief network (DBN) is proposed. Variable importance projection analysis is adopted to screen variables, the time delay of each variable is estimated, and the phase space of the original sample is reconstructed by analyzing the historical data. An ensemble strategy based on random subspace is presented, including the data set partition method and ensemble mode of the model. First, subspaces are constructed according to the component information extracted by partial least squares. Then, the deep belief network is used as a submodel. Finally, a back propagation neural network is developed for model combination. The ensemble deep belief network model has been used to model the NO_x emission prediction of a 660 MW boiler. The simulation results show that the ensemble DBN model can fully exploit the nonlinear mapping relationship between input variables and NO_x concentration by using various learning learners. Compared with the back propagation neural network and support vector machine, which are commonly used in NO_x modeling, the ensemble DBN model has better prediction performance and generalization ability.



1. INTRODUCTION

NO_x emission is one of the main pollutants during the combustion process in coal-fired power plants. It not only causes serious environmental problems but also damages human health. Due to the increasingly stringent emission restrictions, NO_x emission reduction technology for power plants has attracted more and more attention from the industry. At present, flue gas denitrification technology and low NO_x emission combustion technology are usually used to reduce NO_x emission for coal-fired boilers.

Selective catalytic reduction (SCR) is a common denitrification method in power plants, which has the advantages of high efficiency and simple equipment. Through SCR, the injected ammonia and NO_x are mixed in the flue gas, and under the catalytic action of catalysts, a redox reaction is carried out in the reactor to generate nitrogen and water.^{1,2} In this process, the amount of ammonia injected is very important. If the amount of ammonia is too small, the NO_x emission cannot be effectively reduced. Excessive ammonia can reduce NO_x emission, but it will cause waste of ammonia, increase the operating cost of the unit, and even cause ammonia leakage, and its byproducts will affect equipment performance. Therefore, timely and accurate injection of an appropriate amount of ammonia into the SCR equipment according to the NO_x content in the flue gas is important to achieve a low NO_x emission.³

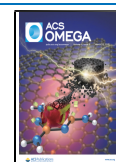
At present, the continuous emission monitoring system (CEMS) is widely used to measure NO_x concentration in denitrification systems. Measurement of NO_x concentration in flue gas must go through the heat pipe and analysis cabinet, which requires a certain amount of time, resulting in a time delay in CEMS measurement. The measured results cannot reflect the change in NO_x concentration at the inlet of the SCR reactor in real time. At the same time, the measurement delay will affect the control of ammonia injection in the subsequent denitrification system, failing feed-forward response, which increases the difficulty of ammonia injection control. Therefore, an accurate NO_x prediction model can be established to predict NO_x emission at the next moment to achieve a redundant measurement with CEMs.

The formation mechanism of NO_x in the furnace is very complex, which involves a variety of chemical reactions and thermal phenomena.⁴ The construction of an ideal mechanism model to predict the dynamics of NO_x emissions is still

Received: December 29, 2020

Accepted: March 1, 2021

Published: March 11, 2021



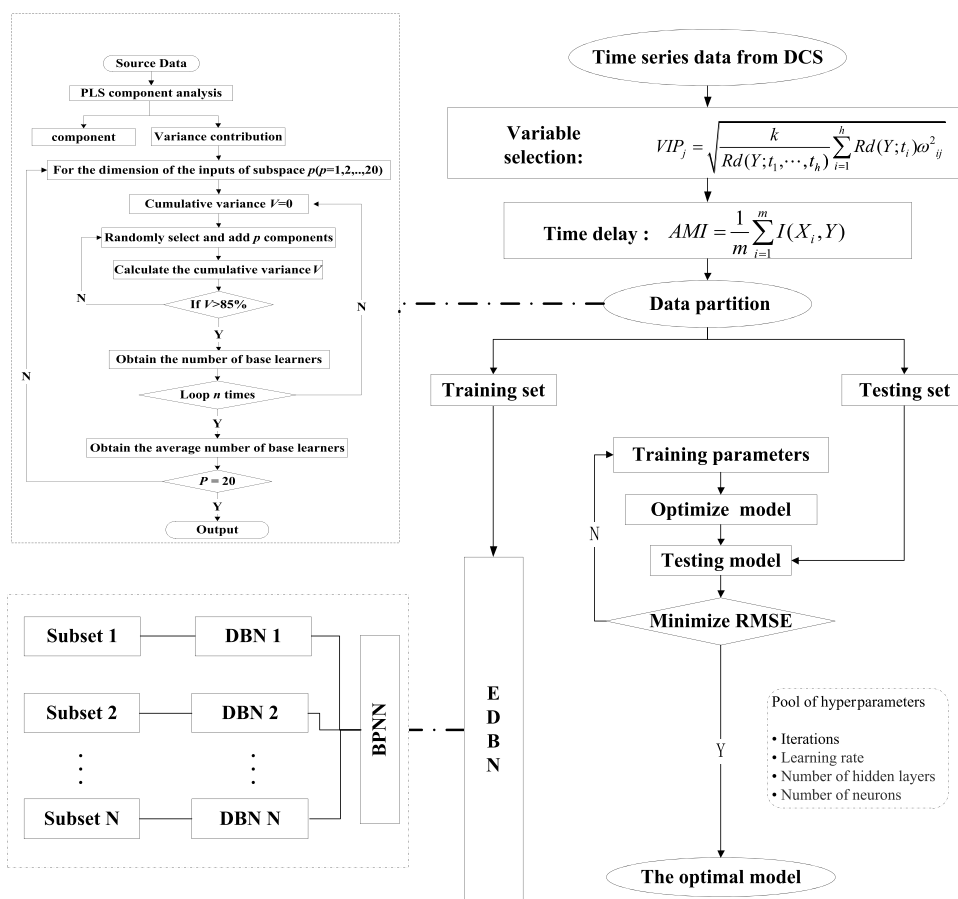


Figure 1. Framework of NO_x emission prediction.

challenging. The data-driven approaches provide a new way to solve this problem: The mapping relationship between boiler operating parameters and NO_x emission is established through offline training, and then the trained model is applied to online NO_x emission prediction.⁵ Researchers have made extensive analyses and improvements to machine-learning models of NO_x emissions, such as an artificial neural network (ANN) and support vector machine (SVM). Ilamathi et al.⁶ developed an ANN model for NO_x emission prediction based on the experimental data from a 210 MW pulverized coal-fired boiler and obtained an optimum level of operating conditions corresponding to low NO_x emission combined with a genetic algorithm approach. Tuttle et al.⁷ mapped the relationship between the operational parameters and NO_x emission through a genetic algorithm-optimized ANN model. Lv et al.⁸ proposed a novel least squares support vector machine (LSSVM)-based ensemble learning to predict NO_x emission for the 660 MW coal-fired boiler. Tan et al.⁹ developed an SVM model combined with principal component analysis (PCA) for NO_x emission prediction, and the data was acquired from a 1000 MW coal-fired power plant to validate the model. However, most of the above algorithms studied are shallow learning, which has limited ability in data mining and generalization ability when dealing with complex problems.¹⁰ The time delay between variables cannot be considered in these models,¹¹ which has a great impact on improving the generalization of the prediction model. Research on NO_x prediction models considering time delay is limited. Zhai et al.⁴ reduced the time delay between input variables through sequential displacement and transfer entropy (TE). TE is a nonparametric measure that estimates the directed

information flow among stochastic processes to detect cause-effect between variables. However, there are some disadvantages such as complex algorithms and large amounts of high-dimensional computation. Moreover, the sequential displacement only uses the TE between a single input variable and NO_x concentration, and the input variable set is not considered as a whole.

In recent years, deep learning has gradually become one of the hottest research fields in machine learning. With the development of deep learning and its superior performance in feature extraction, a few kinds of research have attempted to develop NO_x emission models with deep learning methods. A restricted Boltzmann machine (RBM),¹² long-short term memory neural network,¹³ deep artificial neural network,¹⁴ and deep belief network (DBN)¹⁵ were applied to modeling NO_x concentration. The restricted Boltzmann machine¹⁶ is successfully used as the structural unit of a deep neural network due to its strong expressive ability. At present, the deep belief network, deep Boltzmann machine (DBM), and other models based on the RBM¹⁷ are considered to be the most effective deep learning algorithms. The DBN is a typical representative of deep learning, which can achieve higher accuracy in data modeling. However, there are two main problems in the DBN: (1) The training process of DBN is complex and time-consuming. (2) Sufficient hidden layers achieve a satisfactory effect, but overfitting may occur when the number of hidden layers is too large. In this paper, an ensemble deep belief network (EDBN) based on random subspace (RS) is proposed, which can solve the above problems and be used to construct a real-time NO_x prediction model of a 660 MW coal-fired power plant. First, the variable

importance projection (VIP) is utilized to select the input variables of the model. Then, the delay time between NO_x concentration and the selected variables is determined by the mutual information method. Finally, the EDBN model is applied to predict the NO_x emission through the selected variables. The performance of the proposed model is compared and analyzed. The framework of NO_x emission prediction is shown in Figure 1.

The rest of the paper is organized as follows. The mechanism of EDBN is illustrated in Section 2. Section 3 introduces the observational data. In Section 4, the methods of variable selection and delay time calculation are proposed. Section 5 presents the experiment results and discussion of the NO_x emission model. Finally, the paper is concluded in Section 6.

2. METHODS

2.1. Deep Belief Network. The DBN is a probabilistic generation model, which is stacked with restricted Boltzmann machines (RBMs). RBMs contain a visible layer and a hidden layer, each containing several neurons. The visible and hidden layers connect only between the layers, and there are no connections within each layer. The work of Le Roux and Bengio shows that RBMs can fit any discrete distribution if the number of neurons in the hidden layer is large enough.¹⁸

The depth network formed by constantly stacking RBMs is the DBN. As shown in Figure 2, in the network, the visible layer

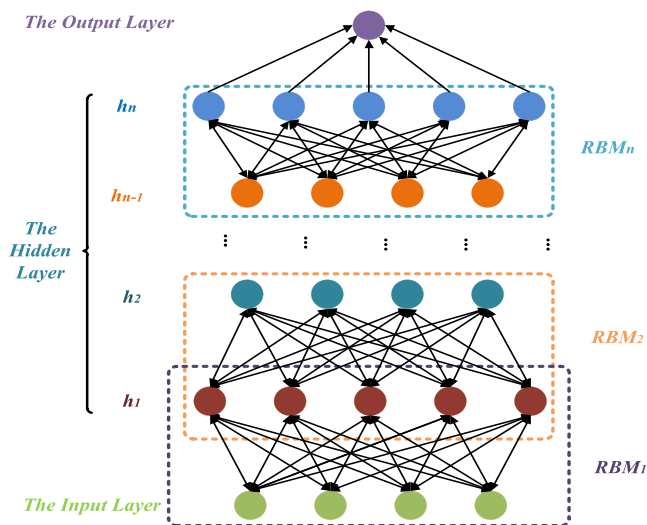


Figure 2. Structure of DBN.

of the first RBM receives input data, and the output of the previous RBM will be the input of the next RBM. The learning process is divided into two stages. First, the unsupervised greedy algorithm is used to train each RBM layer by layer. In this stage, the parameter of each RBM can be obtained. The entire network is then fine-tuned using a supervised back-propagation algorithm. If each RBM contains n neurons in the visible layer and m neurons in the hidden layer, the energy function of the state is

$$E(v, h|\theta) = -\sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i w_{ij} h_j \quad (1)$$

Based on eq 1, the joint probability distribution of (v, h) can be obtained:

$$P(v, h|\theta) = e^{-E(v, h|\theta)} / Z(\theta) \quad (2)$$

$$Z(\theta) = \sum_{v, h} e^{-E(v, h|\theta)} \quad (3)$$

where a and b , respectively, represent the offsets of the visible layer and the hidden layer, v_i denotes the random state of the neurons of the visible layer and h_j denotes the random state of the hidden layer, and w is the weight matrix between the visible layer and the hidden layer. The optimal network parameters can be obtained by training the RBM network $\theta = [w, a, b]$.

The probability of activation of hidden layer neurons h_j is

$$P(h_j = 1|v, \theta) = \sigma \left(b_j + \sum_i v_i w_{ij} \right) \quad (4)$$

Also, the probability of activation of visible layer neurons v_i is

$$P(v_i = 1|h, c) = \sigma \left(a_i + \sum_j w_{ij} h_j \right) \quad (5)$$

where $\sigma(x) = 1/(1 + \exp(-x))$, and the parameters of RBM can be obtained from the logarithmic likelihood of the training set by a gradient descent method, as shown below:

$$\frac{\partial \log p(v)}{\partial \theta} = \sum_{t=1}^N \left[\left\langle \frac{\partial(-E(v^{(t)}, h|\theta))}{\partial \theta} \right\rangle_{p(v^{(t)}, h|\theta)} - \left\langle \frac{\partial(-E(v, h|\theta))}{\partial \theta} \right\rangle_{p(v, h|\theta)} \right] \quad (6)$$

N represents the number of samples in the training set.

2.2. Ensemble Deep Belief Network Algorithm Based On Random Subspace. Although the DBN has a strong ability to express knowledge, it needs to spend a long time to train the model in the fine-tuning stage when processing large-scale data. This takes a lot of training time, and the model is easy to overfit. To solve the above problems, this paper integrates multiple DBN learners and proposes an ensemble DBN model based on random subspace, namely, the EDBN. The random subspace method can construct multiple base learners from the original feature space by randomly selecting the subspace, and one base learner can learn a feature subspace data set.¹⁹ Finally, the output of all the base learners is combined and the final prediction result is obtained through some combination strategies, such as simple average methods and majority voting methods. The structure of EDBN is shown in Figure 3. The data samples are divided, namely, the feature space T is divided into p subspace, and each base learner is trained independently and in parallel on the sample subspace, thus forming p DBN models. The output of EDBN is obtained through a back propagation (BP) neural network.

The performance of the proposed model will be better than the classical DBN prediction algorithm. In this method, the training set is divided into several subsets, and each subset represents the projection of the training set in a subspace. DBN-based learners with the same structure are used for parallel training, and finally, the results are input into the BP neural network for integration. Ensemble learning can accomplish learning tasks by integrating multiple learners, and its performance is better than that of a single learner.

For the random subspace, each randomly selected feature vector can generate a view of the original sample, so multiple

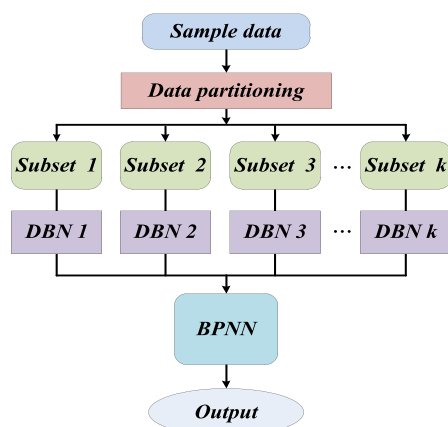


Figure 3. Structure of EDBN.

views will be generated for multiple random sampling features, that is, the sample will be analyzed and described from different views. If the sample has p different representations, there are p views. According to the projection views of the original samples in different subspaces, different base learners with the same structure are designed, and each base learner is trained independently and parallelly. During the training process, the gradient of each base learner's parameter is calculated using the gradient descent update rule, and the calculation method is as follows:

$$\Delta w_{i,j} = \frac{\partial \log p(v)}{\partial w_{i,j}} = \varepsilon (\langle v_j h_j \rangle_{\text{data}} - \langle v_j' h_j' \rangle_{\text{recon}}) \quad (7)$$

$$\Delta a_i = \frac{\partial \log p(v)}{\partial a_i} = \varepsilon (\langle v_i \rangle_{\text{data}} - \langle v_i' \rangle_{\text{recon}}) \quad (8)$$

$$\Delta b_j = \frac{\partial \log p(v)}{\partial b_j} = \varepsilon (\langle h_j \rangle_{\text{data}} - \langle h_j' \rangle_{\text{recon}}) \quad (9)$$

where ε is the learning rate and $\langle \cdot \rangle_{\text{data}}$ and $\langle \cdot \rangle_{\text{recon}}$ denote the expectation of the training sample and the reconstruction model, respectively. The moment gradient descent method is used to modify the parameters of the base learner as follows:

$$(w_{i,j})_n = m(w_{i,j})_{n-1} + \Delta w_{i,j} \quad (10)$$

$$(a_i)_n = m(a_i)_{n-1} + \Delta a_i \quad (11)$$

$$(b_j)_n = m(b_j)_{n-1} + \Delta b_j \quad (12)$$

2.3. Data Partition. With the random subspace method, the scale of the model should be determined first, that is, the dimension of the inputs of the constructed subspace and the number of base learners. To combine base learners with maximum diverse information, these can be obtained through partial least squares (PLS) component analysis²⁰ and a Monte Carlo strategy (Figure 4).

Suppose that the input data $X = [x_1, x_2, \dots, x_A] \in R^{N \times A}$ and the output data $Y \in R^{N \times 1}$, N is the number of sample data and A is the dimension of the input variable. The data are normalized. The input data after PLS feature extraction is $T = [t_1, t_2, \dots, t_A]$, where A is the number of PLS components extracted. The first component t_1 is obtained by a linear combination of x_1, x_2, \dots, x_A , which has the greatest correlation with the input data X and the output data Y . After the regression of the first component t_1 , the residual term of the second component t_2 can be calculated. By

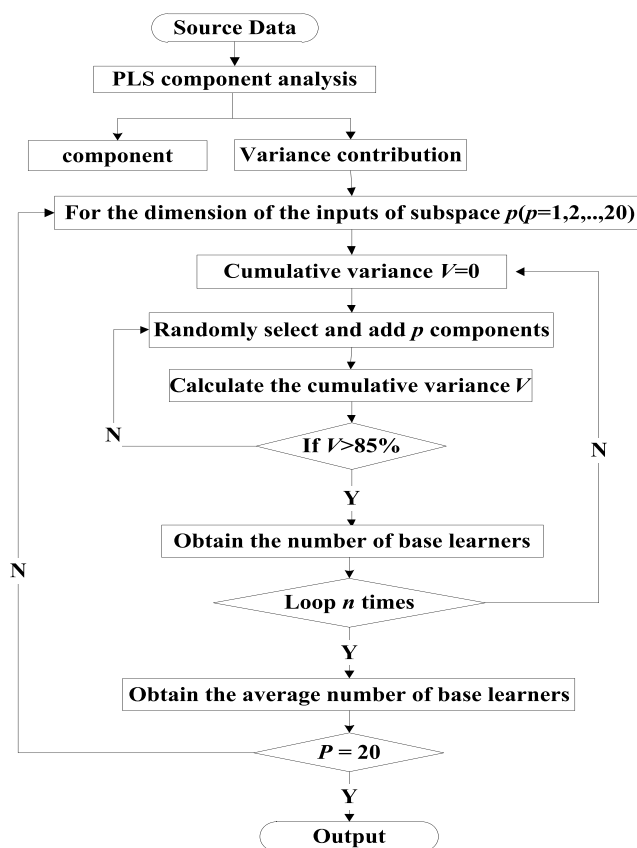


Figure 4. Flowchart of data partitioning.

this method, required PLS components T and corresponding variance Δ captured can be obtained. The process is shown as eqs 13–15

$$X = \sum_i^A t_i p_i^T + E \quad (13)$$

$$Y = \sum_i^A u_i + F \quad (14)$$

$$u_i = b_i t_i + e_i \quad (15)$$

where $P = [p_1, p_2, \dots, p_A]$ is the input load matrix, $W = [w_1, w_2, \dots, w_A]$ is the output weight matrix, $T = [t_1, t_2, \dots, t_A]$ and $U = [u_1, u_2, \dots, u_A]$ are the input and output score matrix, $B = [b_1, b_2, \dots, b_A]$ is the regression coefficient by minimizing the residual, and E and F are the input and output residuals. Additionally, $T = X^* W (P^T W)^{-1}$ and $X^* = \frac{1}{A} \sum_{i=1}^A x_i / \sqrt{\sum_{i=1}^A (x_i - \bar{x})^2 / A}$.

On this basis, the Monte Carlo strategy is used to realize the partition of the subspace. The steps are as follows:

S_1 : Through the PLS algorithm, the original training data set $X = [x_1, x_2, \dots, x_p]$ can be reconstructed as $T = [t_1, t_2, \dots, t_k]$, and the resulting variance contribution matrix $\Delta = [\lambda_1, \lambda_2, \dots, \lambda_k]$.

S_2 : Set the dimension of the inputs of the constructed subspace $p (p = 1, 2, \dots, k)$.

S_3 : Randomly select p components as the input of the base learner and calculate the cumulative variance.

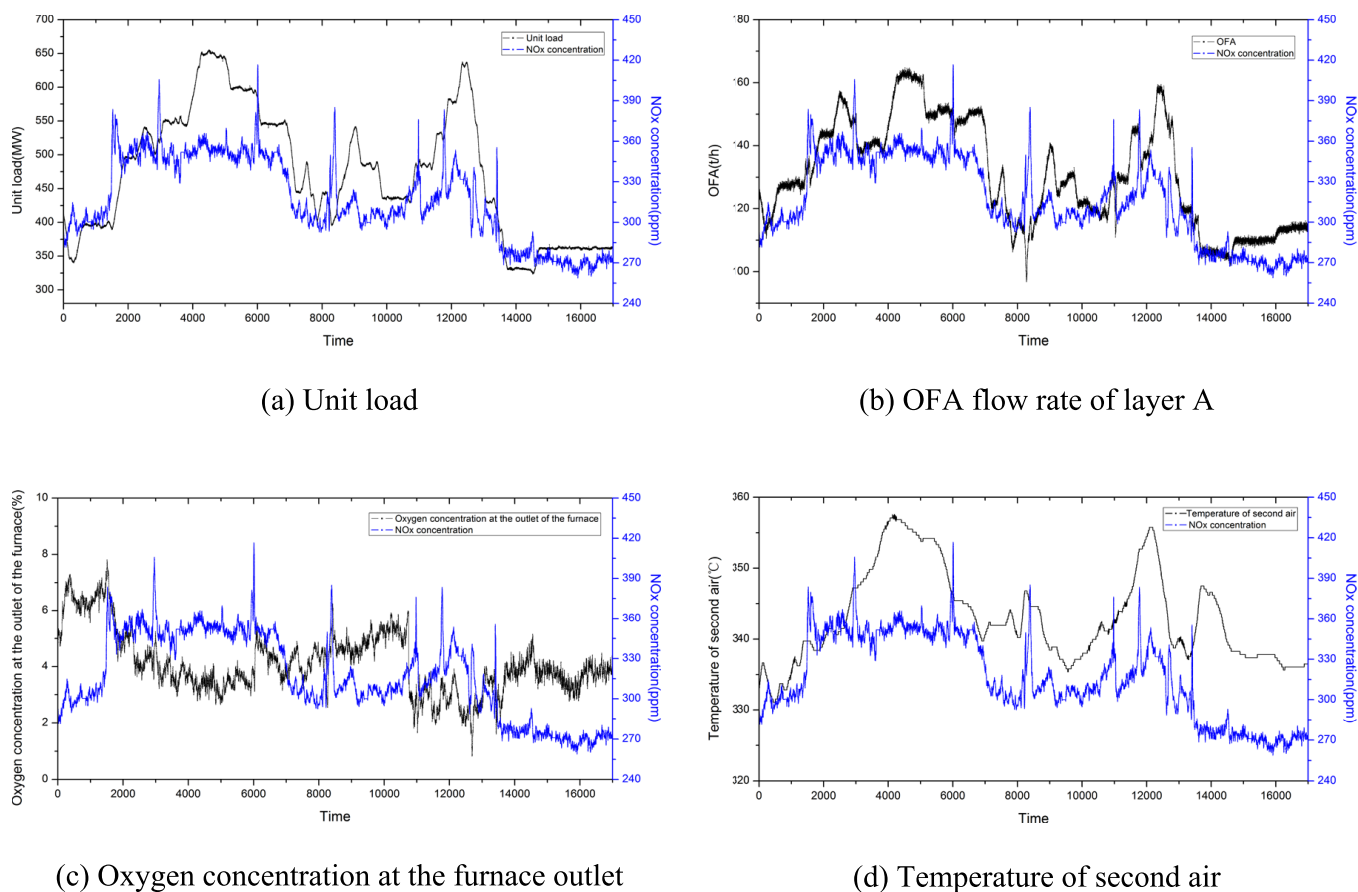


Figure 5. (a–d) Dynamic changes in boiler operating parameters and NO_x concentration.

S_4 : Among the remaining components, continue to select p components as the new subset and calculate the cumulative variance.

S_5 : Repeat S_4 until the cumulative variance of all selected components reaches 85% after q ($q = 1, 2, 3, \dots$) selection.

S_6 : Repeat S_3 – S_5 until n ($n \geq 10000$) iteration; q_i is the number of base learners of the i iteration.

S_7 : Obtain the number of base learners $q^* = \sum q_i/n$.

Thus, the relationship between the dimension of the inputs of the constructed subspace p and the number of base learners q is obtained. The scale of the ensemble model determines the degree of the description of the original data by random subspace. If p or q is too small, the interpretation ability of the ensemble model will be lower and the information contained will be less. If q is too large, the risk of dimensional disaster increases, and the redundancy of the model increases. The sample data is partitioned into q subsets: $X \rightarrow T \rightarrow \{T^1, T^2, \dots, T^q\}$.

2.4. Evaluation Metrics. The following metrics are employed to evaluate the performance of the EDBN model:

$$\delta_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (16)$$

$$\delta_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| \times 100\% \quad (17)$$

$$\delta_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (18)$$

where δ_{MAE} is the mean absolute error (MAE), δ_{MAPE} is the mean absolute percentage error (MAPE), δ_{RMSE} is the root-mean-square error (RMSE), y_i is the real NO_x emission, y'_i is the NO_x emission predicted by the neural network, and n is the total number of test samples.

3. DATA DESCRIPTION

The research object of this paper is the 660 MW coal-fired boiler with ultra-supercritical parameters, which adopts an opposed-wall-firing mode. The data are acquired from the distributed control system (DCS) of the power plant, the boiler load of which varied from 300 to 660 MW with comprehensive data coverage. The sampling interval of the data is 5 s, and more than 10,000 operation data of the boiler are recorded. NO_x produced by the coal-fired boiler is mainly fuel-type NO_x and thermal-type NO_x. There are many factors affecting the formation of NO_x, such as coal quality, air/coal ratio, and temperature in the main combustion zone. In addition, for a given boiler, the combustion operation of the boiler will have a great influence on NO_x emissions. Considering the basic knowledge of the NO_x formation and the suggestions of the engineer, the variables related to the boiler operation and NO_x formation are selected. Due to the lack of an on-line coal analyzer, the real-time coal quality data cannot be obtained, and the kind of coal does not change during the data collection process. More importantly,

Table 1. Variables with VIP Values

unit load	total air rate	coal-feed rate	flue gas temperature at the furnace outlet	primary air temperature	main steam flow rate	over-fire air flow rate			
						A	B	C	D
1.40	1.41	1.39	1.35	1.25	1.39	1.36	0.84	1.39	1.28
main steam pressure	total air primary rate	oxygen concentration at furnace outlet	secondary air temperature	second air flow rate					
				A	B	C	D	E	F
1.41	1.33	1.04	1.09	1.19	1.35	1.13	1.17	0.89	1.19

coal quality can be reflected by operational variables and the historical sequence of NO_x emissions.

The dynamic changes in operating parameters and NO_x concentration of the 660 MW coal-fired boiler are obtained, as shown in Figure 5. Some nonlinear relationships between NO_x concentration and boiler operating parameters can be observed. The unit load, OFA flow rate of layer A, and second air temperature are positively correlated with NO_x emission to some extent. NO_x concentration is highly correlated with unit load, and NO_x concentration when the boiler is under high load is much greater than that when the boiler is under low load. There may be some delay or negative correlation between NO_x concentration and the oxygen concentration at the outlet of the furnace. The NO_x generation mechanism is complex, and it is difficult to establish a mechanism model to describe the nonlinear relationship. In contrast, the data-driven modeling approach does not need to consider complex mechanistic processes, and it establishes a nonlinear model to describe a complex relationship based on input and output data. A large amount of process data generated during boiler operation provides the basis for data-driven modeling.

4. VARIABLE SELECTION AND TIME DELAY

4.1. Variable Selection. Data-driven models are sensitive to data, and the input of the model directly affects its prediction accuracy and generalization ability. Insufficient input variables will lead to inaccurate prediction, but too much input will increase computational complexity and reduce prediction accuracy. NO_x emissions from power stations are affected by a variety of variables. Studies on the formation mechanism of NO_x have revealed the main factors affecting NO_x, but these studies are usually carried out by field tests or numerical calculations. Considering the sensitivity of the neural network to data, to better select variables based on the mechanism research, the data analysis method is used to screen variables.

Variable importance projection (VIP) is a variable screening method based on partial least squares regression. When multiple independent variables have a strong correlation, it describes the explanatory ability of independent variables to dependent variables through the synthesis principal component of the dependent variable and selects independent variables according to their explanatory ability. Chen et al.²¹ pointed out that VIP values reflect not only the importance of independent variables to the model but also the expression of dependent variables. For the data with strong correlation, the VIP method can be used appropriately and accurately to screen the independent variables. Assuming the dependent variable y and the independent variables x_1, x_2, \dots, x_k , the VIP value of the j independent variable to dependent variable y can be expressed as

$$VIP_j = \sqrt{\frac{k}{Rd(Y; t_1, \dots, t_h)} \sum_{i=1}^h Rd(Y; t_i) \omega_{ij}^2} \quad (19)$$

$$Rd(Y; t_1, \dots, t_h) = \sum_{i=1}^h Rd(Y; t_i) \quad (20)$$

$$Rd(Y; t_i) = r^2(Y; t_i) \quad (21)$$

where k denotes the number of independent variables, h denotes the total number of components, ω_{ij} represents the weight value of the i th variable in the j th component, $Rd(Y; t_1, \dots, t_h)$ denotes the explanatory ability of t_1, \dots, t_h to Y , $Rd(Y; t_i)$ is the explanatory ability of t_i to Y , and $r(Y; t_i)$ represents the correlation coefficient.

Due to the multidimension and complexity of the training sample, VIP can be used to extract the data to the maximum extent and continuously extract effective information from the residual so as to obtain an appropriate input data set. On this basis, the multicollinearity between variables can be weakened to a certain extent, and the low-dimensional input data can be used as far as possible to obtain the predicted results. The greater the VIP between the boiler operation parameter and the NO_x emission, the more important the relevant parameter is to the NO_x emission sequence and the more suitable it is to be used as the input of the prediction model. To select variables, VIP values are sorted in descending order. Then, cutoff thresholds can be estimated subjectively based on process knowledge or through iteration to optimize some desired performance criterion. In this paper, independent variables with a VIP value less than 0.8 are considered as low-contribution variables, which can be eliminated. The operation parameter data is processed through VIP variable selection, and the analysis results are shown in Table 1.

4.2. Time Delay. During the production, multiple operational parameters are affecting NO_x emission. However, the measurement of these parameters cannot be obtained instantaneously, that is, the measurement of different parameters has a corresponding delay time. The existence of time lag between measurement data will result in the data that cannot reflect the actual operation at the current moment. Moreover, the size of time delay parameter has a significant influence on the performance of time series prediction;²² the nonlinear relationship between NO_x concentration and boiler parameters cannot be correctly reflected by the established prediction model. Therefore, these time delays need to be determined first to guide subsequent modeling.

Mutual information²³ comes from the concept of entropy in information theory, and as an information measure, it reflects the degree of statistical dependence between two variables. For the industrial process with long time lag, this paper utilizes the mutual information method to estimate the time delay of each input variable of the model.

The information entropy of the random variable X is defined as

$$H(X) = - \sum_{i=1}^N P(x_i) \log_2 P(x_i) \quad (22)$$

where $H(X)$ represents information entropy and $P(X)$ represents the probability distribution for a discrete random variable X by sample size of N that gets values x_1, x_2, \dots, x_N , with probabilities of p_1, p_2, \dots, p_n .

Mutual information between two random variables X and Y is determined by

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \quad (23)$$

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \log_2 \frac{P(x, y)}{P(x)P(y)} \quad (24)$$

where $P(X, Y)$ is the joint distribution of variables X and Y and $P(X)$ and $P(Y)$ are the marginal distributions of X and Y , respectively.

The input variable set can be defined as $X(t) = [X_1(t), X_2(t), \dots, X_m(t)]$; m denotes the number of the input variables. $Y(t)$ is the output, which is the NO_x concentration. Input variables need to be considered as a whole, so the delay time for each variable is calculated based on the average mutual information (AMI) between multiple variables,²⁴ which is defined as

$$\text{AMI} = \frac{1}{m} \sum_{i=1}^m I(X_i, Y) \quad (25)$$

Since the time delay between each input $X_i(t)$ and the output $Y(t)$ is different, phase space reconstruction is performed on each $x_i(t)$, and the input matrix embedded with different time delays $\tau_i \in [\tau_{\min}, \tau_{\max}]$ is obtained $X = [X_1(t - \tau_1), X_2(t - \tau_2), \dots, X_m(t - \tau_m)]$. τ_{\min} and τ_{\max} are the minimum and the maximum possible delay time of input variables, respectively, the values of which are determined by field experience. Considering the actual situation of the unit in the paper and the suggestions of the operators, the time delay ranges from 5 to 300 s, and τ_{\min} and τ_{\max} are set to 1 and 60, respectively.

According to eq 25, the AMI among variables during the different embedding time delay was calculated. When the AMI value is maximum, the corresponding τ is the delay of the input variable.

Considering the number of input variables m and the time range of possible delays $\tau_{\max} - \tau_{\min}$, an exhaustive search to perform this minimization algorithm must explore $m^{\tau_{\max} - \tau_{\min}}$ possible solutions and compute $m^{\tau_{\max} - \tau_{\min}}$ times. The computation of an exhaustive search algorithm makes this method not feasible in practical application. In order to overcome this problem, particle swarm optimization (PSO) is used to jointly estimate the time delay between input variables and the output variable. As a common optimization algorithm, PSO has been widely used in many industrial applications. It can be used to solve complex nonlinear problems with fast computing speed and a wide application range. The parameters are set as follows: the population size is 200, the maximum number of iterations is 100, the acceleration coefficients c_1 and c_2 are both equal to 2 and remain unchanged in the searching process, and the lower and upper bounds of the inertia weight factor ω are 0.4 and 0.9, respectively. The goal is to minimize the value of AMI. The sampling interval of the original data is 5 s. The actual delay time is calculated using the input variables selected in Section 4.2, and the results are shown in Table 2.

Table 2. Time Delay Estimated Results of Input Variables

serial no.	variable name	τ	delay time(s)	input variable	input variable after adjustment
1	unit load	6	30	$x_1(t)$	$x_1(t - 6)$
2	total air rate	40	200	$x_2(t)$	$x_2(t - 40)$
3	coal-feed rate	31	155	$x_3(t)$	$x_3(t - 31)$
4	main steam pressure	34	170	$x_4(t)$	$x_4(t - 34)$
5	main steam flow rate	44	220	$x_5(t)$	$x_5(t - 44)$
6	total air primary rate	49	245	$x_6(t)$	$x_6(t - 49)$
7	OFA flow rate of layer A	30	150	$x_7(t)$	$x_7(t - 40)$
8	OFA flow rate of layer B	41	205	$x_8(t)$	$x_8(t - 41)$
9	OFA flow rate of layer C	37	185	$x_9(t)$	$x_9(t - 37)$
10	OFA flow rate of layer D	35	175	$x_{10}(t)$	$x_{10}(t - 35)$
11	second air flow rate of layer A	34	170	$x_{11}(t)$	$x_{11}(t - 34)$
12	second air flow rate of layer B	46	230	$x_{12}(t)$	$x_{12}(t - 46)$
13	second air flow rate of layer C	56	280	$x_{13}(t)$	$x_{13}(t - 56)$
14	second air flow rate of layer D	37	185	$x_{14}(t)$	$x_{14}(t - 37)$
15	second air flow rate of layer E	28	140	$x_{15}(t)$	$x_{15}(t - 28)$
16	second air flow rate of layer F	35	175	$x_{16}(t)$	$x_{16}(t - 35)$
17	primary air temperature	34	170	$x_{17}(t)$	$x_{17}(t - 34)$
18	secondary air temperature	44	220	$x_{18}(t)$	$x_{18}(t - 44)$
19	flue gas temperature at furnace outlet	10	50	$x_{19}(t)$	$x_{19}(t - 10)$
20	oxygen concentration at furnace outlet	36	180	$x_{20}(t)$	$x_{20}(t - 36)$

5. RESULTS AND DISCUSSION

5.1. Data Partition. As shown in Table 2, the paper adjusts the selected auxiliary variables to the unified timing sequence according to the calculated time delay, and the variables after adjustment will be used for the next work. By PLS component analysis and the Monte Carlo strategy, the adjusted input sequence is partitioned into different subsets. Figure 6 shows the results of PLS component analysis, and Figure 7 shows the relationship between the dimension of components in subspace and the number of base learners.

From Figure 7, it can be seen that the cumulative variance capture increases with the increase in the number of input components of subspace. When the cumulative variance capture is required to be higher than 85%, the dimension of input components and the number of base learners are negatively correlated to a certain extent. When the dimension of components in subspace is greater than six, the descend rate of the number of base learners gradually slows down with the increase in input component dimension. On the premise that the cumulative variance capture meets the requirements, we should follow the principle of keeping the model scale as simple as possible. Therefore, the number of base learners is five and the dimension of components in subspace is six for ensemble learning.

5.2. NO_x Emission Prediction. The original data is reconstructed to reduce the influence of time delay between variables. The reshaped data is divided into the training set and

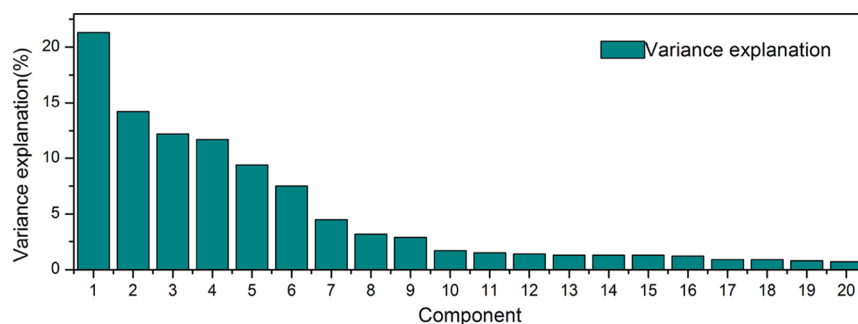


Figure 6. Variance explanation of PLS components.

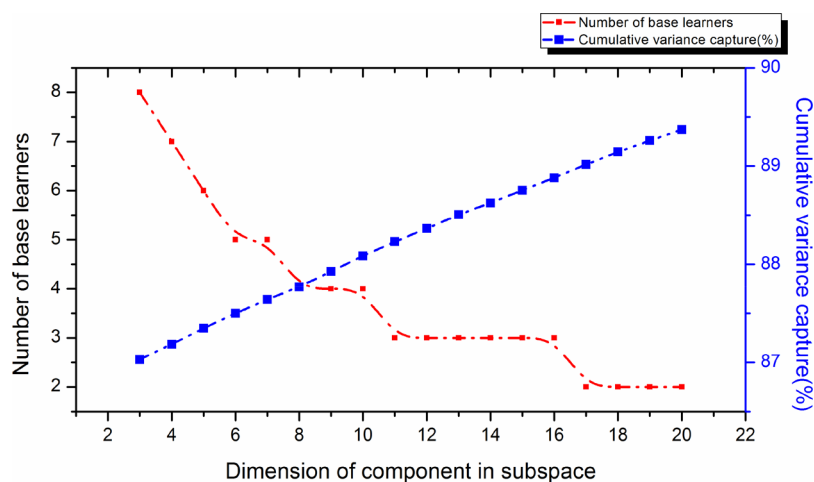


Figure 7. Relationship between the dimension of components in subspace and the number of base learners.

testing set. The hyperparameters of the base learner are listed in Table 3.

Table 3. Optimal Hyperparameters of the Base Learner

hyperparameter	value
iterations	1000
number of hidden layers	2
number of neurons	[100,400]
learning rate	0.001

Figure 8 shows the NO_x emissions between the predicted and measured values of the data set. As a training set, the first 6000 data are training sets used to verify the learning ability of the EDBN model. It can be seen that the predicted values of EDBN are consistent with the measured values. Moreover, when the NO_x concentration changes with time, the predicted values of EDBN can completely track its change trend, indicating that the EDBN has a good learning ability for the original training data.

As a testing set, the last 1000 data are used to verify the generalization ability of the EDBN model. Compared to the prediction results of the training set, the predicted values of the

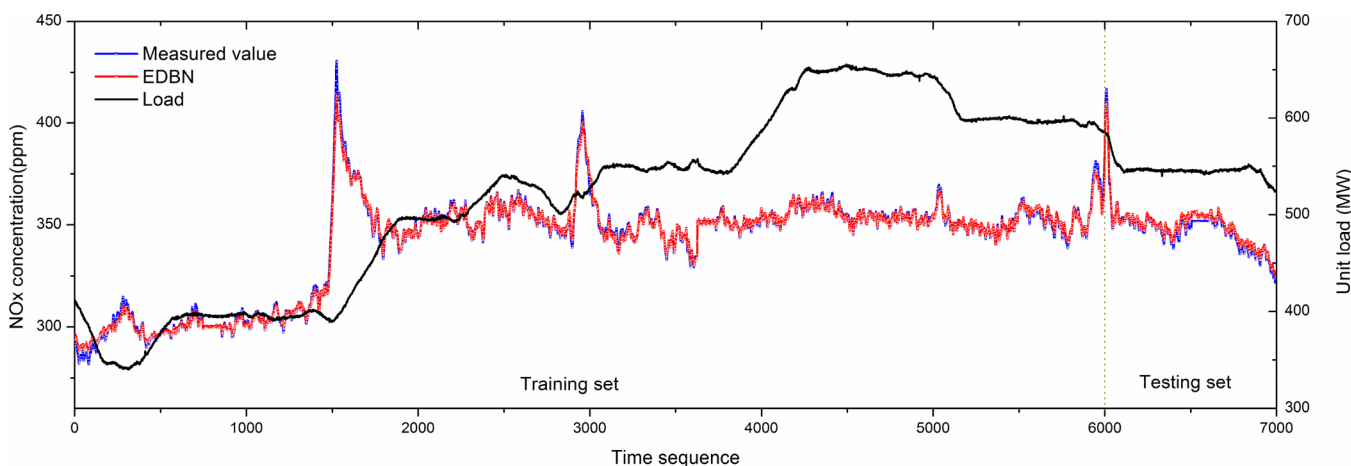


Figure 8. Measured and predicted NO_x concentration by the EDBN model.

testing set fluctuate slightly relative to the measured values. According to the calculations, the EDBN model has high accuracy, where MAPE = 0.566%, MAE = 1.970 ppm, and RMSE = 2.304 ppm. Moreover, while the NO_x concentration changes with time, the predicted values of the EDBN model are well tracked, indicating that the EDBN model has good generalization ability, which can realize the prediction of NO_x concentration.

In the study, the BP neural network is used to integrate the DBN base learners. At the same time, the ensemble strategy of the weighted average is also implemented, which is the ADBN shown in Table 4. The average strategy is widely used in

Table 4. Comparison of EDBN and Base Learner Results

model	RMSE (ppm)		MSE (ppm)		MAPE (%)	
	train	test	train	test	train	test
EDBN	1.979	2.304	1.328	1.970	0.386%	0.566%
ADBN	2.771	3.225	1.860	2.758	0.543%	0.793%
DBN	3.958	4.608	2.657	3.940	0.776%	1.133%
Base1	6.379	7.890	5.161	5.393	1.501%	1.537%
Base2	7.647	7.863	6.277	5.435	1.816%	1.550%
Base3	6.307	7.990	4.877	6.076	1.418%	1.750%
Base4	6.772	7.793	5.415	5.687	1.572%	1.629%
Base5	7.356	8.484	5.270	6.405	1.524%	1.846%

ensemble learning. In addition, the DBN model with all 20 input variables is also established. Also, all models have 5-fold cross-validation.

Of all the learning models applied in Table 4, each base learner performs worst in the prediction of the training set and the testing set. The main reason is that different DBN base learners receive different input components of the same dimension, but they cannot fully contain all the feature information related to NO_x concentration. Therefore, some information may be missing. This results in the base learner that performs even worse than the DBN model with all 20 variables as an input. Moreover, it is difficult to obtain an excellent result by simply averaging the base learners. Better predictive performance depends on the integration strategy. Compared with the base DBN learner and 20-input DBN model, the EDBN exhibits better performance on the testing set. This is the advantage of ensemble learning, which can make up for the different prediction effects of base learners.

For the comparison of ensemble methods, the performance of the BP ensemble is better than that of the average ensemble. BP learning can explore the predictive performance of different base learners, which is equivalent to the adaptive weighted integration of different base learners. However, the performance of the average ensemble is easily affected by the outliers in the base learners, so the prediction result of the average ensemble is weaker than that of the BP ensemble.

In addition, Figure 9 shows the performance of EDBN, ADBN, and DBN models on testing set data in more detail. In the aspect of accuracy, the prediction results of the three models all have a slight fluctuation compared with the measured value of NO_x concentration, but the prediction accuracy of EDBN and ADBN is nearly doubled compared with that of DBN. In terms of variation trend, the predicted values of EDBN or ADBN can fully track the trend of NO_x concentration, and the real-time prediction effect of EDBN is better.

5.3. Comparisons with Other Methods. The BP and SVM are also used in this study to establish comparative models,

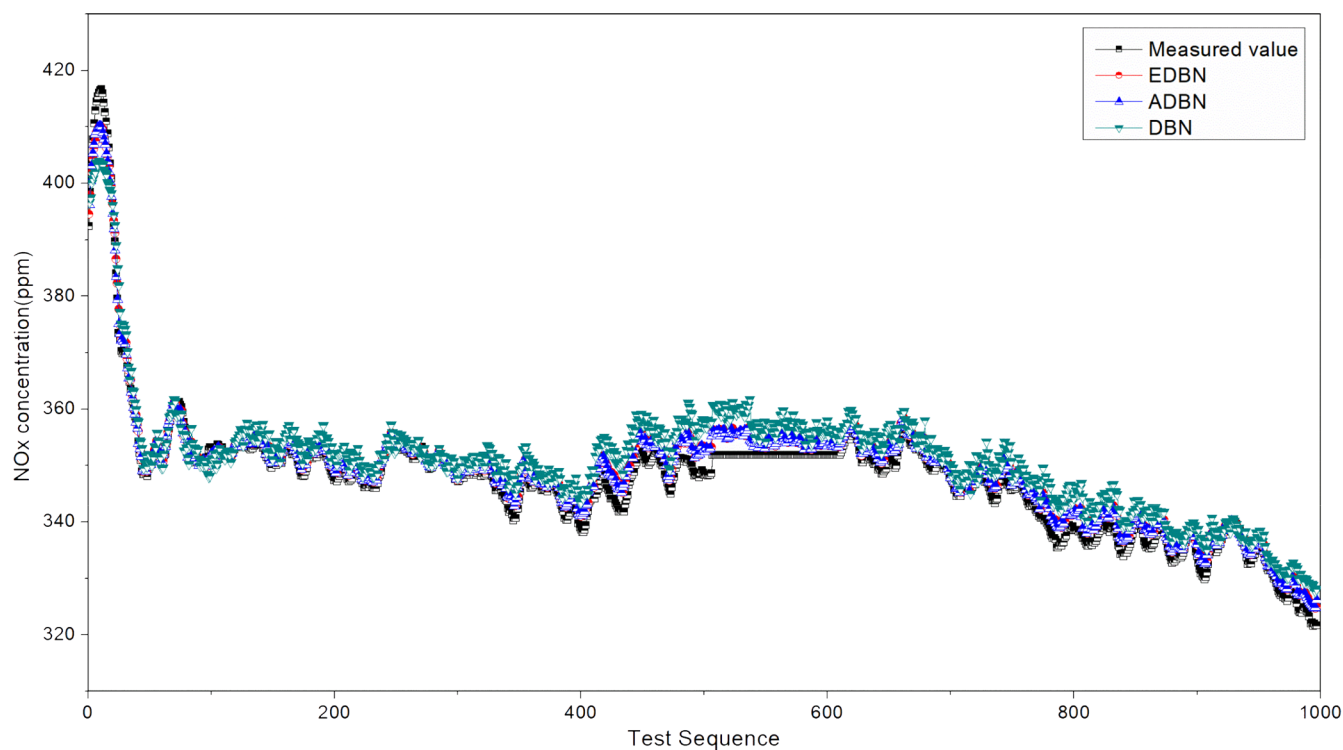
both of which have been successfully and widely used in NO_x emission modeling. The widely used BP neural network is a feed-forward network, which can be considered as nonlinear mapping of the input pattern to the output pattern. A three-layer network with Relu hidden neurons is selected to accomplish the model, and the number of neurons was determined by repeated attempts. The training set and test set are consistent with the EDBN model. The model is constructed by the training set, and the performance of the model is verified by the test set. Similar to most SVM modeling research studies, the generalization performance of the model mainly depends on two parameters, namely, the generalization parameter C and kernel function parameters γ .²⁵ PSO is also used to optimize these two parameters so as to ensure high precision prediction results and optimal generalization performance. Similarly, BP and SVM models are established based on the data adjusted for time delay.

To compare the performance of each model more intuitively, Figure 10 shows the distribution of the estimation errors predicted by the various models on the testing set. It can be found that the three deep learning models show good approximation accuracy, and the prediction errors vary within a small range around [−14, 12]. The BP and SVM model show poor performance, and the maximum prediction error range is around [−60, 30]. The errors generated by the deep learning model are greatly reduced. In particular, the EDBN and ADBN, the two ensemble models, have most errors within 5%, exhibiting better prediction accuracy in general on the testing set, because the prediction errors are distributed more closely to the zero line on the graph than other models. It further proves the advantages of ensemble learning.

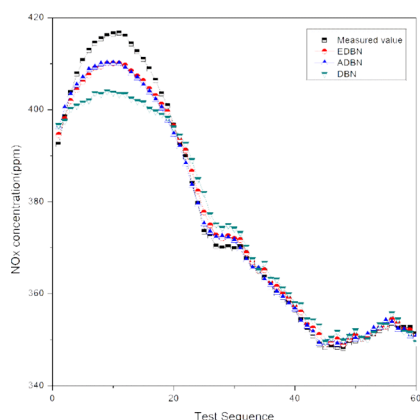
The boiler combustion process and SCR system have the characteristics of large time lag, and the measured values of different measurement points at a time cannot represent the real-time sequence of the process. The reconstruction of time series in Table 2 can eliminate the unnecessary lag and improve the calculation efficiency. The input sequence adjusted is adopted by the models above. In addition, the models of the original sequence as an input are also established: EDBN0, DBN0, SVM0, and BP0.

The impact of VIP variable selection on model prediction accuracy is also compared. The NO_x concentration model is built based on real data to verify the proposed variable selection method. The unit in this paper is similar to that in ref 13. Therefore, for comparison, using the input variables in ref 13, a total of 35 input variables are selected referring to the combustion mechanism. Therefore, 35 parameters are selected as input variables, which include coal-feed rates (A, B, C, D, E, and F), primary air rates (A, B, C, D, E, and F), secondary air rates (A, B, C, D, E, and F), OFA air rates (A, B, C, and D), main steam temperature (1), primary air temperature (2), secondary air temperature (2), main steam flow rate (1), flue gas temperature at the furnace outlet (1), boiler load (1), total air flow rate (1), oxygen concentration at the furnace outlet (1), and furnace temperature of different layers (3). According to the variables selected for the mechanism of NO_x generation, the NO_x concentration models are established: EDBN1, DBN1, SVM1, and BP1.

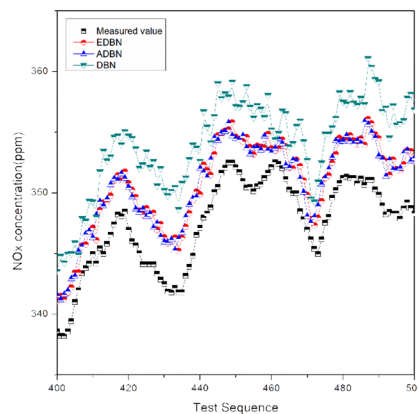
By comparing the boxplot of the estimation error distribution of the testing data from different prediction models, it can be seen that the EDBN and DBN of the deep learning model are much smaller than those of the traditional models BP and SVM. The error distribution of the former is concentrated, which indicates that the deep learning model has a better performance



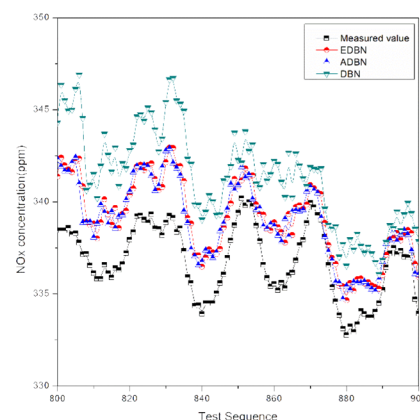
(a) Comparison chart



(b) Declining data segment



(c) Rising data segment



(d) Fluctuating data segment

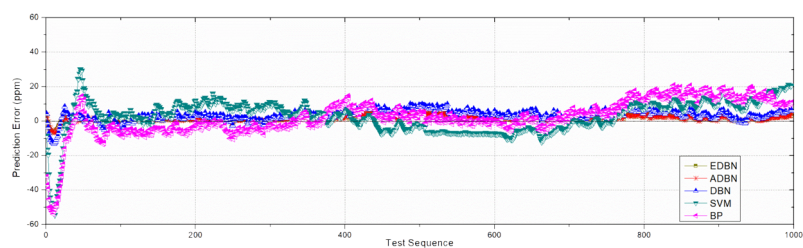
Figure 9. (a–d) Real-time prediction results of NO_x concentration of different models on the testing data set.

in predicting NO_x concentration of thermal power plants. For the same model, the prediction results of the model considering time delay are slightly better than those of the original model, and the median of prediction error is smaller, which indicates that the performance of the model considering time delay is better than that of the original model. At the same time, Table 5 makes a quantitative analysis of the performance indexes of each model, and their visual comparison is shown in Figure 11. It can be found that the prediction accuracy of the models after sequence reconstruction is higher. This indicates that the delay time should be taken into account when establishing an accurate dynamic model.

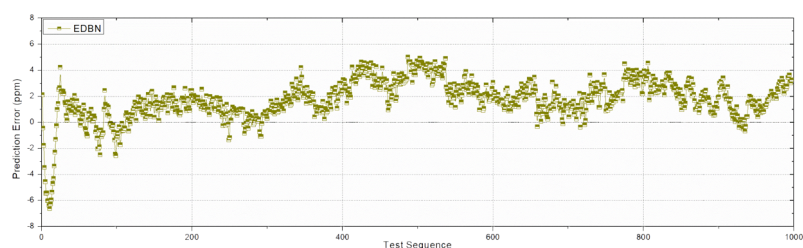
In addition, when establishing the NO_x emission prediction model of thermal power units, the prediction error of the model with VIP analysis is smaller than that with variables selected

according to mechanism analysis. If the NO_x emission model is directly established according to the mechanism analysis, although the forecast trend of the model is still the same as the original data, its prediction error is much higher than that of the prediction model after selecting the input variables. This indicates that after VIP variable selection, a more effective set of input variables is obtained, the number of input variables is reduced, and the prediction accuracy is improved while reducing the complexity of the model and showing better generalization ability.

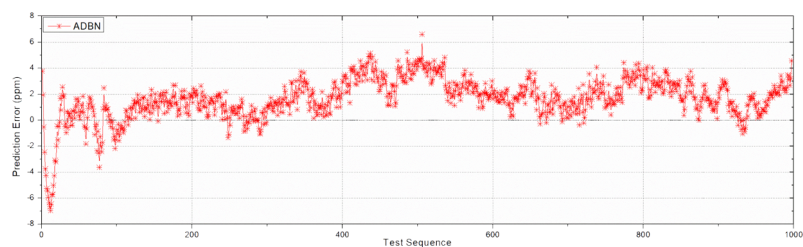
In conclusion, the prediction based on the method presented in this paper has better prediction performance. The EDBN can better track the change trend of NO_x concentration value. It shows that the EDBN has a good ability to learn data, which reflects the advantages of the ensemble model. In terms of



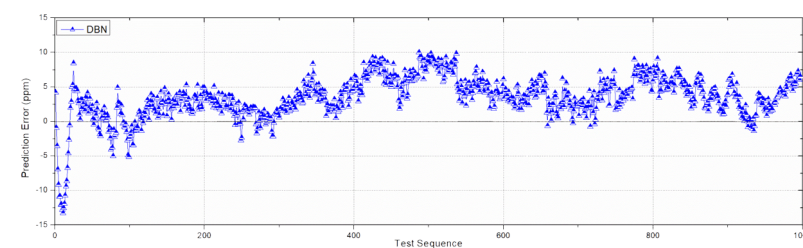
(a) Comparison chart



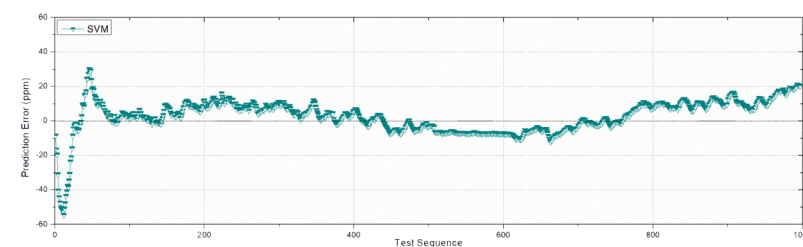
(b) EDBN



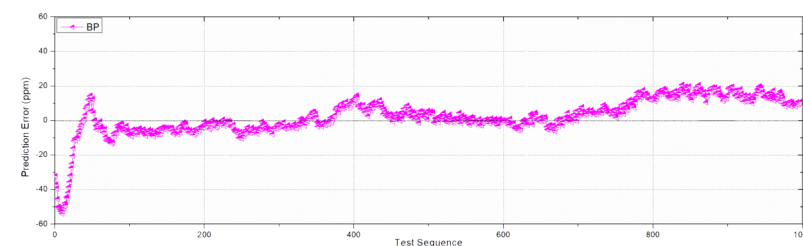
(c) ADBN



(d) DBN



(e) SVM



(f) BPNN

Figure 10. (a–f) Prediction errors of the testing data for various models.

Table 5. Performance Comparison of Different Models

model	performance index		
	RMSE	MAPE (%)	MAE
EDBN			
35-inputs (mechanism analysis)	5.831	1.532	5.326
20-inputs (VIP analysis)	4.147	1.020	3.546
20-inputs-reconstruction (considering time delays)	2.304	0.566	1.970
DBN			
35-inputs (mechanism analysis)	10.533	2.782	9.668
20-inputs (VIP analysis)	5.529	1.359	4.728
20-inputs-reconstruction (considering time delays)	4.608	1.133	3.940
BPNN			
35-inputs (mechanism analysis)	17.743	4.273	14.838
20-inputs (VIP analysis)	13.593	2.85	9.880
20-inputs-reconstruction (considering time delays)	10.795	2.15	7.525
SVM			
35-inputs (mechanism analysis)	17.081	4.386	15.356
20-inputs (VIP analysis)	12.792	3.12	10.981
20-inputs-reconstruction (considering time delays)	10.153	2.18	7.644

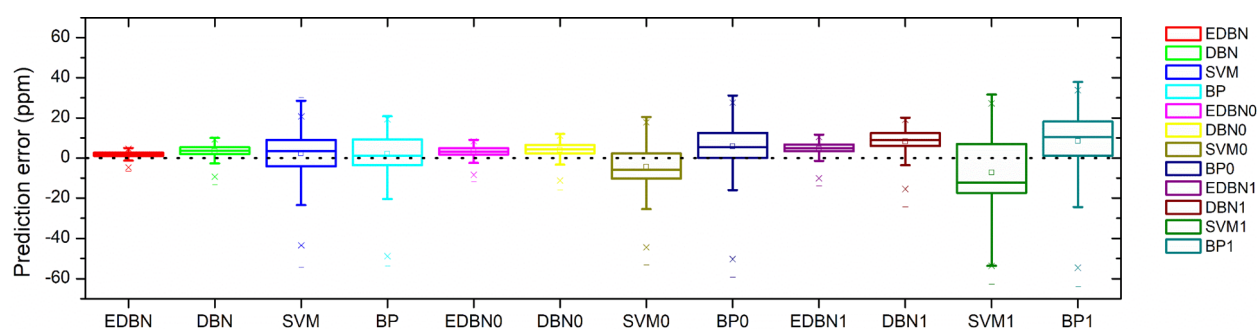
sequence, the reconstruction of sequence plays an important role in improving the prediction performance. Therefore, adjustment sequence is very important to model prediction and cannot be ignored. The time delay is detected by statistical analysis of the data without understanding the mechanism of the system. The selection of input variables is also necessary. Insufficient input variables will lead to inaccurate prediction, but

too much input will increase computational complexity and reduce prediction accuracy. Scientific variable screening has been proven to be an effective method to improve the accuracy of model prediction. When selecting variables in practical application, we should pay attention to important explanatory factors and follow the principle of keeping variables as few as possible.

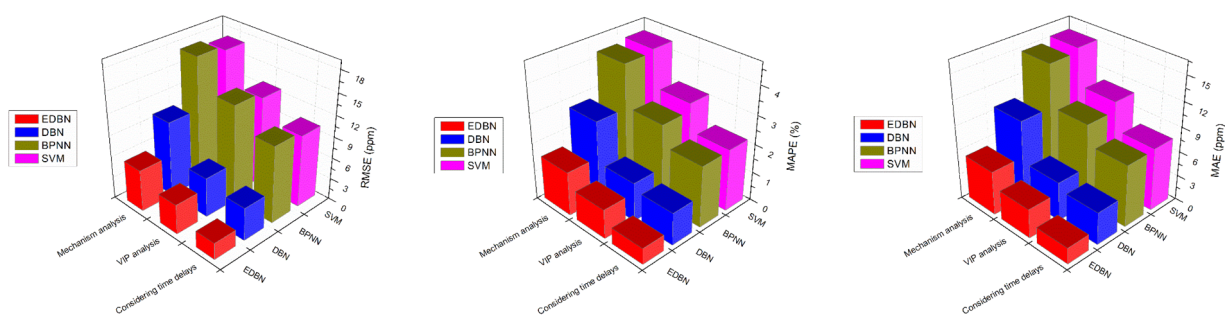
6. CONCLUSIONS

The establishment of an effective NO_x prediction model is the basis for reducing NO_x emissions. In this study, the EDBN model has been successfully established to predict the NO_x emissions of a 660 MW ultra-supercritical coal-fired power plant using historical operating data. The major conclusions are as follows:

- (1) The data-driven model is sensitive to data, and the input of the model directly affects its prediction accuracy and generalization ability. To better select variables, based on mechanism research, the VIP analysis method is used to screen variables.
- (2) There is a delay between the measurement of operating variables and NO_x concentration at the SCR inlet in the furnace, which can be accurately described quantitatively by analyzing historical data. The delay time between input variables and NO_x concentration is calculated based on AMI, and the PSO algorithm is used to estimate the delay time.
- (3) An ensemble strategy based on random subspace is proposed, including the data set partition method and ensemble mode of model. The sample data is divided according to the component information extracted by



(a) Boxplot of prediction error



(b) RMSE

(c) MAPE

(d) MAE

Figure 11. (a–d) Prediction error distribution of the testing data for various models.

PLS, and the sample subspaces are constructed. Then, DBN base learners are trained in each sample subspace, and finally, the BP network is applied to obtain the result of ensemble model.

- (4) The ensemble DBN model has been used to model the NO_x emissions prediction. The ensemble DBN model can take advantage of each base learner and fully explore the nonlinear mapping relationship between input characteristics and NO_x concentration so as to improve the prediction accuracy of the ensemble model. Compared with the BP and SVM, which are commonly used in NO_x modeling, the EDBN model has better prediction performance. This is mainly due to the limited capacity of shallow networks in processing large data sets.
- (5) The phase space reconstruction of samples is carried out by estimating the time delay of each input variable. Based on this, the NO_x emission model is established by the data rearranged according to the delay time. By comparing the models before and after data reconstruction, the prediction results show that the model after data reconstruction obtains better performance for predicting NO_x emission.

AUTHOR INFORMATION

Corresponding Author

Yingnan Wang – School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China; Email: yingnan@ncepu.edu.cn

Authors

Guotian Yang – School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Ruibiao Xie – School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China; orcid.org/0000-0002-2297-8478

Han Liu – School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Kai Liu – School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Xinli Li – School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsoomega.0c06317>

Funding

This research was funded by the Fundamental Research Funds for the Central Universities, China (grant no. 2018QN052).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the friendly support, supply of design data, operational measurements, and technical advice of Heqi Power Plant and the financial support of the Fundamental Research Funds for the Central Universities, China (2018QN052).

ABBREVIATIONS

AMI, average mutual information; ANN, artificial neural network; BMCR, boiler maximum continue rate; BP, back propagation; CEMS, continuous emission monitoring system;

DBM, deep Boltzmann machine; DBN, deep belief network; DCS, distributed control system; EDBN, ensemble deep belief network; GA, genetic algorithm; LSSVM, least squares support vector machine; MAE, mean absolute error; MAPE, mean absolute percentage error; OFA, over fire air; PCA, principal component analysis; PLS, partial least squares; PSO, particle swarm optimization; RBM, restricted Boltzmann machine; RBMs, restricted Boltzmann machines; RMSE, root-mean-square error; RS, random subspace; SCR, selective catalytic reduction; SVM, support vector machine; VIP, variable importance projection

REFERENCES

- (1) Colombo, M.; Nova, I.; Tronconi, E. Detailed kinetic modeling of the NH₃-NO/NO₂ SCR reactions over a commercial Cu-zeolite catalyst for Diesel exhausts after treatment. *Catal. Today* **2012**, *197*, 243–255.
- (2) Schobing, J.; Tschamber, V.; Brilhac, J.-F.; Auclair, A.; Hohl, Y. Simultaneous soot combustion and NO_x reduction over a vanadia-based selective catalytic reduction catalyst. *C. R. Chim.* **2018**, *21*, 221–231.
- (3) Xie, P.; Gao, M.; Zhang, H.; Niu, Y.; Wang, X. Dynamic modeling for NO_x emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network. *Energy* **2020**, *190*, 116482.
- (4) Zhai, Y.; Ding, X.; Jin, X.; Zhao, L. Adaptive LSSVM based iterative prediction method for NO_x concentration prediction in coal-fired power plant considering system delay. *Appl. Soft Comput.* **2020**, *89*, 106070.
- (5) Wang, C.; Liu, Y.; Zheng, S.; Jiang, A. Optimizing combustion of coal fired boilers for reducing NO_x emission using Gaussian Process. *Energy* **2018**, *153*, 149–158.
- (6) Ilamathi, P.; Selladurai, V.; Balamurugan, K.; Sathyanathan, V. T. ANN-GA approach for predictive modeling and optimization of NO_x emission in a tangentially fired boiler. *Clean Technol. Environ. Policy* **2013**, *15*, 125–131.
- (7) Tuttle, J. F.; Vesel, R.; Alagarsamy, S.; Blackburn, L. D.; Powell, K. Sustainable NO_x emission reduction at a coal-fired power station through the use of online neural network modeling and particle swarm optimization. *Control Eng. Pract.* **2019**, *93*, 104167.
- (8) Lv, Y.; Liu, J.; Yang, T.; Zeng, D. A novel least squares support vector machine ensemble model for NO_x emission prediction of a coal-fired boiler. *Energy* **2013**, *55*, 319–329.
- (9) Tan, P.; Zhang, C.; Xia, J.; Fang, Q.; Chen, G. NO_x emission model for coal-fired boilers using principle component analysis and support vector regression. *J. Chem. Eng. Jpn.* **2018**, *49*, 211–216.
- (10) Bengio, Y. *Learning deep architectures for AI*; Now Publishers Inc.: 2009, DOI: 10.1561/9781601982957.
- (11) Wu, X.; Zhu, X.; Wu, G.-Q.; Ding, W. Data mining with big data. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 97–107.
- (12) Fan, W.; Si, F.; Ren, S.; Yu, C.; Cui, Y.; Wang, P. Integration of continuous restricted Boltzmann machine and SVR in NO_x emissions prediction of a tangential firing boiler. *Chemometr. Intell. Lab. Syst.* **2019**, *195*, 103870.
- (13) Yang, G.; Wang, Y.; Li, X. Prediction of the NO_x emissions from thermal power plant using long-short term memory neural network. *Energy* **2020**, *192*, 116597.
- (14) Adams, D.; Oh, D.-H.; Kim, D.-W.; Lee, C.-H.; Oh, M. Prediction of SO_x-NO_x emission from a coal-fired CFB power plant with machine learning: Plant data learned by deep neural network and least square support vector machine. *J. Cleaner Prod.* **2020**, *270*, 122310.
- (15) Wang, F.; Ma, S.; Wang, H.; Li, Y.; Zhang, J. Prediction of NO_x emission for coal-fired boilers based on deep belief network. *Control Eng. Pract.* **2018**, *80*, 26–35.
- (16) Hinton, G. E. A practical guide to training restricted Boltzmann machines. In *Neural networks: Tricks of the trade*; Springer: 2012; pp. 599–619.

- (17) Lee, S.; Chang, J.-H. Deep belief networks ensemble for blood pressure estimation. *IEEE access* **2017**, *5*, 9962–9972.
- (18) Le Roux, N.; Bengio, Y. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput.* **2008**, *20*, 1631–1649.
- (19) Deng, Z.; Jiang, Y.; Chung, F.-L.; Ishibuchi, H.; Choi, K.-S.; Wang, S. Transfer prototype-based fuzzy clustering. *IEEE Trans. Fuzzy Syst.* **2015**, *24*, 1210–1232.
- (20) Baffi, G.; Martin, E. B.; Morris, A. J. Non-linear projection to latent structures revisited: the quadratic PLS algorithm. *Comput. Chem. Eng.* **1999**, *23*, 395–411.
- (21) Chen, X.; Huang, J.; Yi, M. Cost estimation for general aviation aircrafts using regression models and variable importance in projection analysis. *J. Cleaner Prod.* **2020**, *256*, 120648.
- (22) Abbasimehr, H.; Shabani, M.; Yousefi, M. An optimized model using LSTM network for demand forecasting. *Comput. Ind. Eng.* **2020**, *106435*.
- (23) Ludwig, O., Jr.; Nunes, U.; Araújo, R.; Schnitman, L.; Lepikson, H. A. Applications of information theory, genetic algorithms, and neural models to predict oil flow. *Commun. Nonlinear Sci. Numer. Simul.* **2009**, *14*, 2870–2885.
- (24) Yang, T.; Ma, K.; Lv, Y.; Bai, Y. Real-time dynamic prediction model of NO_x emission of coal-fired boilers under variable load conditions. *Fuel* **2020**, *274*, 117811.
- (25) Zhou, H.; Zhao, J. P.; Zheng, L. G.; Wang, C. L.; Cen, K. F. Modeling NO_x emissions from coal-fired utility boilers using support vector regression with ant colony optimization. *Eng. Appl. Artif. Intell.* **2012**, *25*, 147–158.