China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 11, No. 3, September 2024 https://doi.org/10.15878/j.instr.202400189

Article

A NO*x* Concentration Prediction Model Based on a Sparse Regularization Stochastic Configuration Network

Aijun Yan[1,2,3], Shenci Cao[1,2]

[1] School of Information Science and Technology, Beijing University of Technology, Beijing 100124, China;

[2] Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China;

[3] Beijing Laboratory for Urban Mass Transit, Beijing 100124, China

[*] Corresponding author email: yanaijun@bjut.edu.cn

Abstract: For accurate prediction of nitrogen oxides (NO*x*) concentration during the municipal solid waste incineration (MSWI) process, in this paper, a prediction modeling method based on a sparse regularization stochastic configuration network is proposed. The method combines DropConnect regularization with L_1 regularization. Based on the L_1 regularization constraint stochastic configuration network output weights, DropConnect regularization is applied to the input weights to introduce sparsity. A probability decay strategy based on network residuals is designed to address situations where the DropConnect fixed drop probability affects model convergence. Finally, the generated sparse stochastic configuration network is used to establish the model, and is validated through experiments with standard datasets and actual data from an MSWI plant in Beijing. The experimental results prove that this modeling method exhibits high-precision prediction and generalization ability while effectively simplifying the model structure, which enables accurate prediction of NO*x* concentration.

Keywords: municipal solid waste incineration; NO*x* concentration prediction; stochastic configuration network; sparse regularization

Citation: Yan, Aijun, and Shenci Cao. "A NO*x* Concentration Prediction Model Based on a Sparse Regularization Stochastic Configuration Network." Instrumentation 11, no. 3 (September 2024): 189. https://doi.org/10.15878/j.instr.202400189.

0 Introduction

In step with economic growth and the urbanization process, a large amount of municipal solid waste (MSW) is generated, leading to increasingly serious environmental pollution. Solid waste incineration has become the primary MSW treatment method in China because of its advantages of polution reduction, resource utilization and harmlessness[1]. However, the nitrogen oxide (NO*x*) formed during municipal solid waste incineration (MSWI) is considered a major atmospheric pollutant, and if the emission concentration exceeds the standards, it can lead to serious health and environmental problems[2]. Currently, continuous emission monitoring systems are widely used to measure NO*x* concentration[3], however, monitoring systems are costly to install and maintain, and the measurement results lag. Therefore, accurate prediction of the NO*x* concentration is crucial for controlling pollution emissions from the MSWI process.

In recent years, NO*x* prediction modeling has benefited from the rapid development of intelligent technologies. Common prediction models are generally divided into mechanistic models and data-driven models. Owing to the complex physical and chemical reactions during NO*x* generation and emission, constructing mechanistic models requires a considerable number of parameters to set up partial differential equations, which makes it difficult to ensure the accuracy and reliability of these models[4]. Therefore, data-driven modeling methods have gradually become an important means to predict NO*x* emissions, especially with the use of the neural network-based modeling methods, which have been widely applied[5-7]. Currently, neural networks mainly include gradient-based neural networks that use the backpropagation (BP) algorithm and randomized learning

neural networks. Among them, the randomized learning neural network overcomes the issues of slow convergence and the local minima trap tendencies of the traditional gradient-based neural networks, and demonstrates fast and effective modeling capabilities[8]. For example, the random vector functional link network[9] simplifies the learning process and improves learning efficiency by randomly assigning input weights and biases of the hidden layer nodes and using the least squares method to calculate output weights. However, the general approximation of such a randomized learning neural network depends on the range of random parameters and the number of hidden layer nodes, which affects its generalizability and learning ability[10,11]. In response to this shortcoming, a stochastic configuration network (SCN) with inequality constraints was presented in [12]. The SCN uses the incremental learning approach to construct the hidden layer, node-by-node, incorporates a supervision mechanism to select random parameters via inequality constraints, and then determines the output weights via the least squares method. Thus, the SCN ensures a general approximation and improves the precision and efficiency of the network. Owing to these advantages, the SCN has been intensively researched and widely used in various areas, including for fault diagnosis[13], parameter prediction[14,15], and modeling of industrial processees[16,17].

In practical applications of neural network modeling, overfitting easily occurs because of the limited training data and complex network structure. To address this issue, regularization strategies are commonly employed to improve the generalization performance of the model. Common regularization methods include L_1 and L_2 regularization, which introduce the norm penalty term into the objective function to constrain the weights. L_2 regularization shrinks the parameters to zero, obtaining a smoother solution, but it does not reduce the complexity of the network. In contrast, L_1 regularization tends to set some parameters to zero, promoting sparsity[18]. In addition, Dropout[19] and DropConnect[20] regularization are also effective methods for preventing overfitting. Dropout breaks the co-adaptation between nodes by randomly dropping some nodes during network training to reduce network overfitting. DropConnect, as an extension of Dropout, randomly drops out connection weights and introduces dynamic sparsity to the connection weights during training[21]. Overfitting also occurs in the SCN's modeling process. With the increase in the number of hidden layer nodes, redundant nodes are inevitably generated because of the characteristics of the randomized algorithm, which complicates the network structure and results in poor robustness and generalization performances. Furthermore, in the application of practical industrial systems, it is often necessary to reduce the model complexity to alleviate the storage pressure while meeting the accuracy requirements. In this context, a parsimonious SCN with L_1 regularization was proposed in [22], which introduced L_1 regularization into the SCN to obtain sparse output weights and improved inequality constraints. Experiments proved that this method can effectively simplify the model structure and improve the generalization performance. In [23], Lu proposed a sparse SCN based on Bayesian learning, assuming that the output weights obey the Laplace distribution, adopting a two-level hierarchical prior distribution as the lower bound of the sparse prior, and obtaining an approximate Gaussian posterior, which simplifies parameter solving while ensuring sparsity, and effectively improves the model accuracy and generalizability ability. It has been proved that the sparse regularization method can simplify the model structure and alleviate the overfitting problem at the same time.

However, owing to the random configuration of the input weights of the SCN, current research on sparse SCN has focused mainly on output weights. This results in the inability to fully exploit the performance of the sparse SCN. Therefore, this paper proposes a sparse regularization stochastic configuration network (SR-SCN) that combines DropConnect and L_1 regularization. On the basis of the L_1 regularization SCN, we use an improved DropConnect regularization method with a probability adaptive decay strategy based on network residuals to introduce sparsity into the input weights of the SCN, breaking the co-adaptation between nodes. SR-SCN has been used to predict the NO*x* concentration in the MSWI process, and the effectiveness of the method was verified via simulation experiments.

The rest of the paper is arranged as follows. Section 1 outlines the MSWI process and NO*x* formation and elimination. Section 2 describes the SR-SCN algorithm. Section 3 presents an experimental evaluation of this method, and Section 4 drawns a conclusion.

1 Analysis of MSWI Process

This study takes an MSWI plant in Beijing as an example, and Fig.1 depicts the flow diagram of the MSWI process. As shown, the mixed MSW is placed in a solid waste storage tank, where it undergoes fermentation and preliminary drying before being transferred to the combustor. The MSW is subsequently conveyed through the grate, and under the combined action of primary air and high furnace temperature, it undergoes drying, combustion, and burnout processes, transforming into heat, residues, dust, and flue gas. Additionally, a significant amount of heat generated during the incineration process is converted into steam in the waste heat boiler system. Through the steam power generation system, the steam drives the turbine to generate electricity. Finally, in the flue gas cleaning system, flue gas containing various pollutants undergoes purification treatment. In summary, MSWI involves a multitude of complex and unpredictable reactions and exhibits dynamic nonlinearity, multivariable coupling, and other characteristics. Hence, modeling research on such a complex dynamic system as MSWI is extremely challenging.

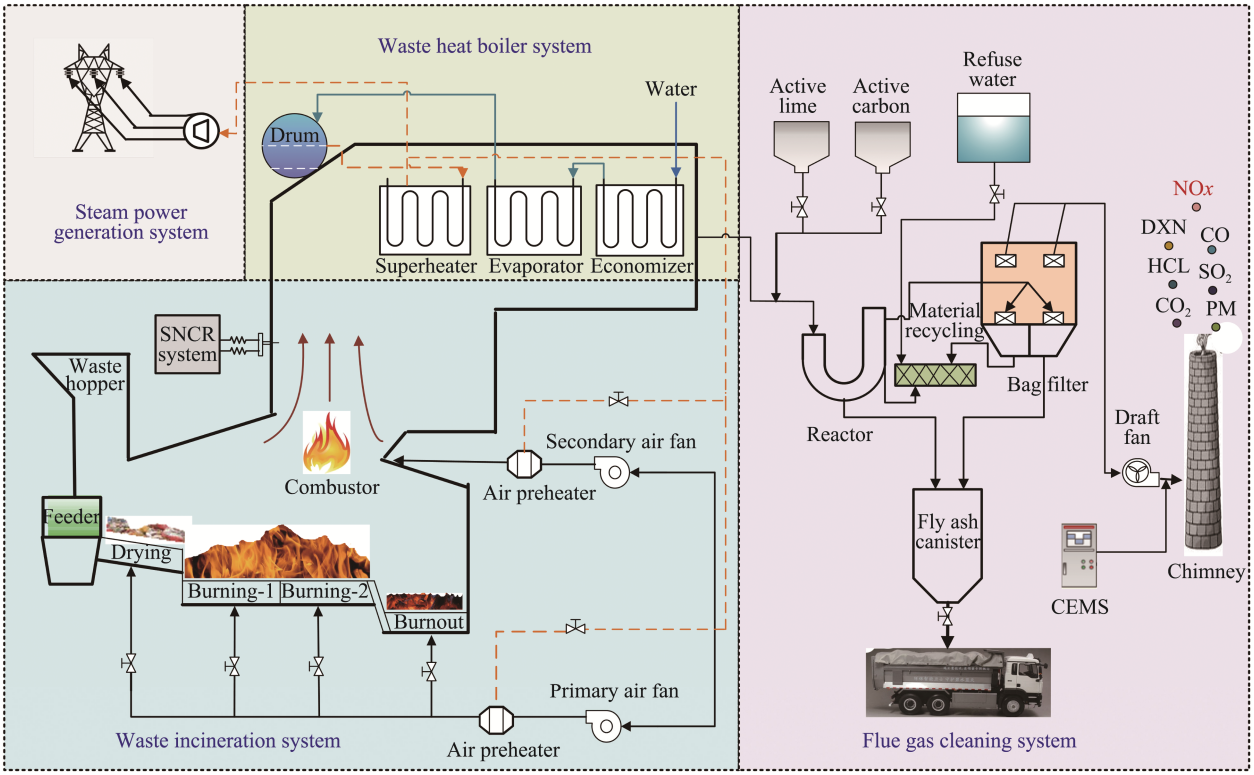

Fig.1 MSWI process flowchart

NOx is one of the important pollutants in the flue gas of the MSWI process and mainly originates from the chemical reaction between the nitrogen compounds of MSW and oxygen, as well as the oxidation of N_2 in the primary air and secondary air at high temperatures. To reduce the generation of pollutants such as NOx and achieve safe and compliant emission of flue gas. Owing to the stable and sufficient combustion of MSW, urea diluent is injected into the furnace through the selective noncatalytic reduction (SNCR) system to eliminate NOx. Finally, the remaining NOx and other flue gases are purified in the flue gas cleaning system through reactors, material recovery, and bag filters before being discharged into the atmosphere. The emission of NOx is influenced by many factors, such as the MSW composition, furnace temperature, air flow rate, and the denitrification and purification processes. Therefore, it is difficult to establish an accurate model to predict NOx emissions from the MSWI process.

2 SR-SCN Algorithm

This section describes the implementation steps of SR-SCN and presents the pseudocode of this algorithm.

2.1 Stochastic Configuration Network

SCN is an incremental randomized learning model that utilizes a supervisory mechanism to randomly assign the input weights and biases of hidden layer nodes; it gradually increases the number of hidden nodes, and uses the least squares method to calculate the output weights. Given a set of training data, with input $X=\{x_1, x_2,\ldots, x_N\}$, output $Y=\{y_1, y_2, \ldots, y_N\}$. For the objective function $f:R^d\rightarrow R^m$, assuming that the SCN has established $L-1$ hidden nodes, the output of the current network is represented as:

$$f_{L-1}(X)=\sum_{k=1}^{L-1}\beta_k g_k(X,\omega_k,b_k) \tag{1}$$

where β_k is the output weight of the hidden layer node k; $g(\cdot)$ denotes the activation function; and ω_k and b_k are the input weights and biases of the k-th node, respectively.

At this point, the network residual is $e_{L\text{-}1}=f-f_{L\text{-}1}$. If the current network residual does not reach the preset error ε, or the maximum number of nodes $L_{\max}$, the L-th hidden layer node is added, and the input weights and biases corresponding to the maximum ξ value for the new node are selected based on the supervisory mechanism of Eq. (3).

$$h_L=[g_L(x_1,\omega_L,b_L),g_L(x_2,\omega_L,b_L),\ldots,g_L(x_N,\omega_L,b_L)]^{\mathrm{T}} \tag{2}$$

$$\xi_L=\frac{\left(e_{L-1}^{\mathrm{T}}\cdot h_L\right)^2}{h_L^{\mathrm{T}}h_L}-(1-r-\mu_L)\,\|e_{L-1}\|^2\geqslant 0 \tag{3}$$

where $h_L\in R^N$ is the output of node L; ω_L and b_L are candidate parameters of node L, and there are a total of $T_{\max}$ sets of candidate parameters; $0<r<1$; and $\{\mu_L\}$ is a sequence of nonnegative real numbers, $\mu_L\leqslant(1-r)$ and $\lim_{L\rightarrow\infty}\mu_L=0$.

Finally, the output weights β are solved according to Eq. (4).

$$\beta=\arg\min_{\beta}\|H_L\beta-Y\|^2=H_L^{\dagger}\mathrm{Y} \tag{4}$$

where $H_L = [h_1, h_2, ..., h_L]$ is the hidden layer output matrix, and $(\cdot)^{\dagger}$ is the Moore-Penrose generalized inverse.

2.2 SR-SCN

DropConnect is a regularization method generalized from Dropout. Unlike Dropout, which drops out hidden layer nodes, DropConnect drops out each connection in the fully connected layer with a certain probability. When DropConnect is applied to the input weights, it is similar to the random subspace method[24]. A subset of features is randomly selected during the training process, introducing randomness to improve the generalization performance of the model.

When applying DropConnect to a fully connected layer, a binary mask M is first drawn, and each element of the mask M is drawn in a *Bernoulli(p)* distribution. The mask is subsequently used to drop out the connections. The probability of a connection being dropped is 1-p, and the probability of being retained is p. A different mask is generated in each iteration. The hidden layer output of applying DropConnect can be expressed as:

$$H = g(X,(M * \omega),b) \tag{5}$$

where each element of the mask M_{ij}~*Bernoulli(p)*.

In this work, DropConnect is applied to the hidden layer input weights of the SCN to obtain a stochastic configuration network with sparse input connections (DC-SCN). When a new hidden layer node L is constructed, the candidate input weights ω_L are randomly dropped via the mask M_L. At this point, the output h_L of node L is rewritten from Eq. (2) as:

$$h_L = [g_L(x_1,\tilde{\omega}_L,b_L), g_L(x_2,\tilde{\omega}_L,b_L),\ldots,g_L(x_N,\tilde{\omega}_L,b_L)]^{\mathrm{T}} \tag{6}$$

where $\tilde{\omega}_L = M_L * \omega_L$. Then, the optimal input weights and biases are selected for the added node according to the supervisory mechanism of Eq. (3).

Owing to the characteristics of SCN incremental learning, in the early stage of model construction, the number of nodes is small, and this model behaves as an underfitting. If too many connections are dropped, the model cannot obtain enough features, which will slow down the convergence of the model. However, if connections are always dropped with a small probability, the complexity of the model cannot be effectively controlled with an increasing number of nodes, and the regularization effect will be seriously affected. Therefore, this paper improves the traditional DropConnect with fixed drop probability, and designs an adaptive decay strategy for probability p based on the network residuals, which is expressed as:

$$p(e_{L-1}) = (1-c) * \exp(-\eta / \| e_{L-1} \|^2) + c \tag{7}$$

where e_{L-1} is the network residual, c is the lower limit of decay, 1-c denotes the range of probability p decay, and η is an adjustable parameter. With increasing network nodes, the network residual gradually decreases, the probability p decreases, and the corresponding connection loss probability gradually increases to ensure the convergence speed and regularization effect of the model.

Based on DC-SCN, L_1 regularization is introduced to establish a sparse constraint on the output weights, and redundant nodes are eliminated to obtain a sparse regularization stochastic configuration network (SR-SCN), as shown in Fig.2. SR-SCN further enhances the generalization ability of the model while simplifying the model structure. By introducing the L_1 norm penalty term into the SCN algorithm, updating the output weights can be simplified to solve the following matrix form:

$$J(\beta) = \lambda_1 \| \beta \|_1 + \frac{1}{2} \| H_L \beta - Y \|^2 \tag{8}$$

The above optimization problem is solved by constructing the augmented Lagrange function and adopting the alternating direction method of multipliers (ADMM). The specific derivation process is described in [22]. The output weight formula is iteratively solved as follows:

$$z^{n+1} = (H_L^{\mathrm{T}} H_L + \rho I)^{-1} (H_L^{\mathrm{T}} Y + \rho(\beta^n - \mu_1^n)) \tag{9}$$

$$\beta^{n+1} = S_{\lambda_1/\rho}(z^{n+1} + \mu_1^n) \tag{10}$$

$$\mu_1^{n+1} = \mu_1^n + z^{n+1} - \beta^{n+1} \tag{11}$$

where λ_1 is the regularization coefficient, ρ is the penalty coefficient, n is the number of iterations, μ_1 is the introduced scale dual variable, S represents the soft threshold operator, and

$$S_{\varphi(a)} = \begin{cases} a-\varphi, a > \varphi \\ 0, |a| \leqslant \varphi \\ a+\varphi, a < -\varphi \end{cases} \tag{12}$$

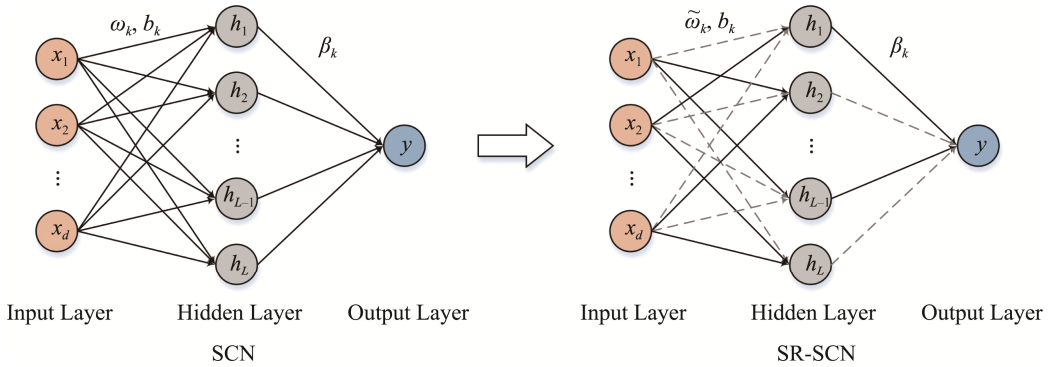

Fig.2 SCN and SR-SCN models

SR-SCN builds upon SCN by introducing sparsity into its input weights via DropConnect and obtaining sparse output weights via L_1 regularization, thereby constructing an overall sparse network structure. Initially, new nodes are created in the SCN, and during the random allocation of input weights, a masking technique is employed to achieve sparsity. The sparsity of the mask is determined by a probability decay function based on the current network residuals. Then, as in SCN, the input weights that satisfy the supervisory mechanism are selected, which ensures the convergence of the algorithm. Finally, when solving the output weights, the L_1-paradigm penalty term is added, and the sparse output weights are solved determined iteratively via the ADMM algorithm. In the process of network construction, the sparse input weights allow the nodes to learn different feature subsets and introduce randomness to break the co-adaptation among nodes. The sparse output weights, on the other hand, play the role of eliminating redundant nodes. Therefore, SR-SCN can effectively simplify the model structure and improve the generalization ability. The pseudocode of the SR-SCN algorithm is as follows:

Algorithm: Sparse regularization stochastic configuration network

Input: Dataset $\{X,Y\}$, $X=\{x_1, x_2,\ldots, x_N\}$, $Y=\{y_1, y_2, \ldots, y_N\}$, the hidden layer maximum node $L_{\max}$, the tolerance error ε, the maximum number of configurations $T_{\max}$, the hidden layer parameter interval Υ, the decay function parameters c and η, the regularization parameters λ_1, ADMM penalty parameter ρ, the maximum number of iterations n_{max};

Output: SR-SCN model.

Initialize: $e_0=[y_1,y_2\ldots,y_N]^{\mathrm{T}}$, $\Omega=[]$, $W=[]$;
While $L \le L_{max}$ *and* $\|e_0\| > \varepsilon$ **do**
  Randomly generate the weights ω_L , bias b_L in Υ;
  Calculate p based on Eq.(7) and generate mask M_L ;
  Calculate h_L, ξ_L based on Eq.(6) and Eq.(3) and save $\tilde{\omega}_L$, b_L in W, save ξ_L in Ω respectively;
  Find $\tilde{\omega}_L^*$, b_L^* that maximize ξ_L in Ω and set $H_L = [h_1^*, h_2^*, \ldots, h_L^*]$;
  Calculate the output weights β^* by Eq.(9)-Eq.(11);
  Calculate $e_L = H_L\beta^* - Y$;
  Renew $e_0 := e_L$; $L := L + 1$;
End While
Return: SR-SCN model

3 Experiments and Results Analysis

In this section, the effectiveness of the SR-SCN method is first validated through experiments using four standard datasets. The SR-SCN method is subsequently used for modeling the prediction of NO*x* concentration in the MSWI process. The simulation experiments utilize historical data from an MSWI plant in Beijing. All the experiments were conducted in the MATLAB R2020b environment.

To validate the performance advantages of the proposed method, SR-SCN, DC-SCN, SCN, and L_1 regularization SCN (L_1-SCN) were compared. The results are the average of 30 independent experiments. The root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were selected as evaluation indices. The formulas are as follows:

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{15}$$

where N is the sample size; y_i is the true value; $\hat{y}_i$ is the predicted value; and $\bar{y}_i$ is the mean true value of the i-th sample, respectively.

3.1 Experiments with Standard Datasets

Experiments were conducted to validate the effectiveness of the proposed method via four single-objective regression standard datasets from the Knowledge Extraction Based on Evolutionary Learning (KEEL) database. For each dataset, the input and output data are normalized to the range of [0, 1]. The activation function of each model is a sigmoid function; the maximum configuration $T_{\max}$=200; the tolerance error $\varepsilon=10^{-4}$; the hidden layer parameter interval Υ=[0.5:0.5:10]; the penalty coefficient ρ=0.05; and the maximum number of iterations n_{max}=150 for the ADMM algorithm. Details of the datasets and the rest of the parameter settings are shown in Table 1.

The experimental results on four standard datasets are shown in Table 2, indicating that the DC-SCN

Table 1 Dataset Information and Parameter Settings

Dataset	Inputs	Training samples	Testing samples	L_{max}	c	η	λ_1
Auto MPG6	5	314	78	50	0.3	10	0.01
Compactiv	21	6144	2048	100	0.1	3	0.01
Concrete	8	618	412	150	0.2	11	0.001
Machine CPU	6	167	42	20	0.7	0.8	0.001

Table 2 Comparison Results of the Algorithms on Standard Datasets

Dataset	Indicator	Algorithm			
		SCN	L_1-SCN	DC-SCN	SR-SCN
Auto MPG6	RMSE	0.0940	0.0685	0.0737	**0.0674**
	MAE	0.0620	0.0457	0.0544	**0.0455**
	R^2	0.8004	0.8945	0.8774	**0.8977**
Compactiv	RMSE	0.0267	0.0228	0.0236	**0.0194**
	MAE	0.0112	0.0131	**0.0106**	0.0128
	R^2	0.9875	0.9912	0.9901	**0.9937**
Concrete	RMSE	0.0883	0.0774	0.0708	**0.0684**
	MAE	0.0598	0.0575	**0.0509**	0.0511
	R^2	0.8164	0.8599	0.8825	**0.8906**
Machine CPU	RMSE	0.1381	0.0407	0.1036	**0.0372**
	MAE	0.0542	0.0242	0.0441	**0.0230**
	R^2	0.1731	0.9365	0.5557	**0.9488**

algorithm outperforms the SCN algorithm across all four datasets in terms of the RMSE, MAE, and R^2 metrics, and yields the smallest MAE in the experiments on the Compactiv and Concrete datasets. It can be demonstrated that DC-SCN has a higher generalization performance than does the SCN. SR-SCN achieves the optimal results on the remaining datasets and metrics with the minimum test errors. Compared with SCN, DC-SCN, and L_1-SCN, SR-SCN has higher test accuracy and generalizability, which proves the effectiveness of the proposed method.

3.2 NOx Concentration Prediction Experiments

3.2.1 Data Preprocessing and Parameter Setting

As the MSWI process is complex and variable, there are many variables affecting NO*x* emission concentration. Therefore, this paper first performs an initial screening of many variables based on the analysis of the MSWI mechanism and then screens the input features of the prediction model based on the minimal-redundancy-maximal-relevance method[25]. The mRMR method aims to maximize the correlation between the features and the target variable and minimize the redundancy among the selected features. In the feature selection process, the relevance is measured by the mutual information between the features and the target, and the redundancy is measured by the mean value of the mutual information between the candidate features and the subset of already selected features. After a feature with the highest relevance is selected as the initial feature subset, the remaining features are incrementally searched based on Eq. (16).

$$\max_{x_j \in X - S_{m-1}} \left[I(x_j;t) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j;x_i) \right] \tag{16}$$

where X is the full feature set, S_{m-1} is the selected feature subset, $I(x_j;t)$ denotes the mutual information between the feature to be selected x_j and the target variable t, and $I(x_j;x_i)$ denotes the mutual information between the feature to be selected x_j and the selected feature x_i.

Based on experimental validation, the impact of the number of selected features on modeling accuracy is presented in Fig.3. The figure shows that the model error is minimized when the number of selected features reaches 20. Furthermore, as the number of features continues to increase, it is difficult to continue improving the accuracy of the model, and contrarily, more computational resources are consumed. Therefore, 20 feature variables are selected as inputs for the predictive model, as shown in Table 3.

The performance of the SR-SCN model was experimentally verified using 1000 historical data from a solid waste incineration plant in Beijing. There were 800 training samples and 200 testing samples. All the data were normalized and linearly mapped to the range of [0,1]. The experimental parameters were set as follows: the maximum number of hidden layer nodes L_{max}=300; in SR-SCN and DC-SCN, the p decay function parameters c=0.1, η=7.2; and in SR-SCN and L_1-SCN, the regularization coefficient λ_1=0.001, while the other parameters remained consistent with those used in the standard dataset experiments.

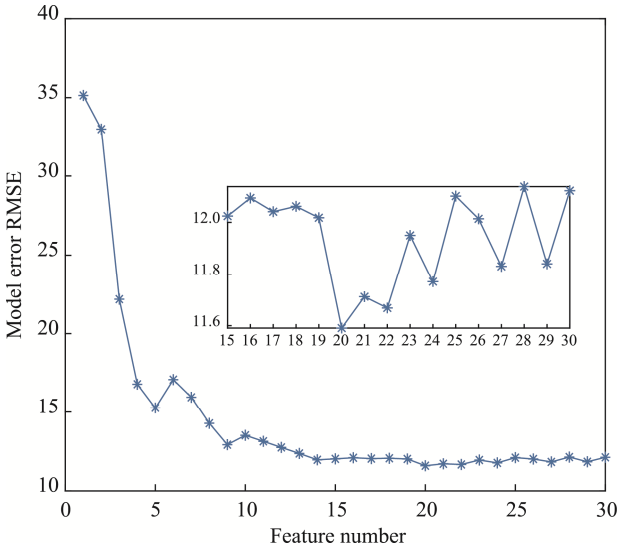

Fig.3 Model Errors with Different Numbers of Features

Table 3 The Input Feature Variables

Index	Variable
1	Flue gas temperature of primary combustion chamber middle 1 (℃)
2	Flue gas temperature of primary combustion chamber right 1 (℃)
3	Flue gas temperature of primary combustion chamber right 3 (℃)
4	Air flow of dry grate left side 1 (Km^3N/h)
5	Air flow of combustion grate left side 1-1 (Km^3N/h)
6	Primary combustion chamber right side temperature (℃)
7	Primary air flow (Km^3N/h)
8	Main steam flow at boiler outlet (t/h)
9	Accumulation of Lime (kg)
10	Accumulation of activated carbon (kg)
11	Flue gas temperature of primary combustion chamber left 2 (℃)
12	Flue gas temperature of primary combustion chamber right 2 (℃)
13	Gas pressure at the outlet of the induced draft fan (Pa)
14	Air flow of dry grate right side 1 (Km^3N/h)
15	Air flow of combustion grate right side 1-1 (Km^3N/h)
16	Primary combustion chamber left side temperature (℃)
17	Secondary air flow (Km^3N/h)
18	Urea solvent flow (L/h)
19	Furnace negative pressure (Pa)
20	O_2 concentration of inlet flue gas (mg/m^3N)

3.2.2 Experimental Results and Analysis

Fig.4 shows the prediction error results of each model under different hidden layer nodes. The optimal prediction errors that can be achieved by each model with the maximum number of nodes in the hidden layer not exceeding 500 are 10.7657 for SCN, 9.4261 for DC-SCN, 9.5434 for L_1-SCN, and 8.5885 for SR-SCN. The prediction accuracy of DC-SCN is improved by 12.44% compared with that of SCN. The prediction accuracy of SR-SCN is improved by 20.22% and 10.01% compared with that of SCN and L_1-SCN, respectively. Fig.4 shows that with a small number of hidden layer nodes, each model exhibits a similar performance. However, as the number of nodes increases, the SCN first experiences overfitting, and the prediction error increases significantly. The DC-SCN has higher prediction accuracy and a better generalization ability than the SCN model, and even within a certain number of nodes, it still has an advantage over the L_1-SCN and SR-SCN models. However, as the number of nodes continues to increase, DC-SCN also tends to overfit. This is because, although DC-SCN simplifies the network structure to a certain extent through sparse input weights, it cannot fundamentally address the complexity of the model caused by increasing nodes. In contrast, SR-SCN introduces L_1 regularization based on DC-SCN, which eliminates the redundant nodes and further simplifies the network model. Compared with the L_1-SCN model, the SR-SCN model has greater prediction accuracy and generalizability. As the number of hidden nodes increases, the advantages of SR-SCN over other models in terms of prediction accuracy and prevention of overfitting become more obvious.

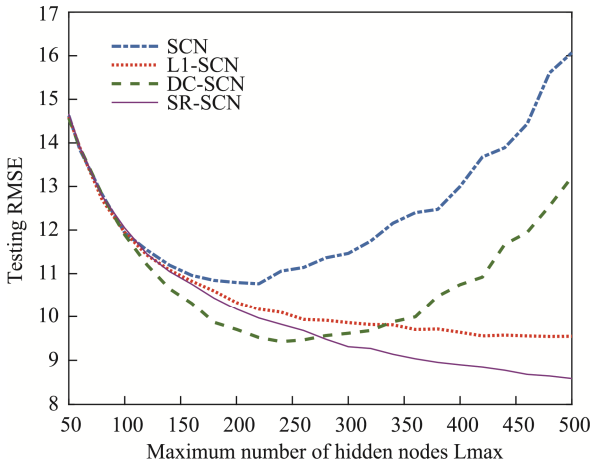

Fig.4 Prediction errors of each model under different hidden layer nodes

Table 4 shows the comparative results of the sparsity of each model when the maximum node number L_{max}=300. The sparsity of each model is measured by the proportion of nonzero elements in the input and output weights. The table shows that the input weights of SR-SCN and DC-SCN are sparse, and their sparsity is nearly identical. The output weights of SR-SCN and L_1-SCN are sparse, and the sparsity of SR-SCN is better than that of L_1-SCN. SR-SCN combines the sparsity of both types of weights, which makes it more effective in simplifying the network structure than the other three methods.

Table 4 Comparison of the Sparsity of Different Models

Method	Sparsity	
	Input weights	Output weights
SCN	1	1
L_1-SCN	1	0.6841
DC-SCN	0.1893	1
SR-SCN	0.1877	0.6404

Table 5 and Fig.5 show the prediction errors and fitting curves of different models, respectively, when the maximum node number L_{max}=300. Table 5 shows that the L_1-SCN, DC-SCN, and SR-SCN models are significantly better than the SCN models in terms of prediction performance. Among the RMSE, MAE, and R^2 metrics, SR-SCN has the best results, with values of 9.2507, 6.9851, and 0.9532, respectively, in terms of the mean values, and the DC-SCN results follow. In the standard deviation (std) comparison, SR-SCN has the best results, with values of 0.2910, 0.2314, and 0.0030, which are followed by the L_1-SCN results. SR-SCN is optimal in the comparison of the means and standard deviations of the three metrics, indicating that SR-SCN has the smallest prediction error, the best fitting effect, and more stable prediction results. As also observed from the fitting curve graph, the SCN model is affected by the overfitting phenomenon and there are obvious errors in the prediction of some samples, whereas the SR-SCN model was fitted with high accuracy and no obvious errors. In summary, the SR-SCN model has the best ability to predict NO*x* emission concentration in the MSWI process.

Table 5 Comparison of Prediction Results of Different Models

Method	RMSE		MAE		R^2	
	mean	std	mean	std	mean	std
SCN	11.6662	0.7232	8.4690	0.4150	0.9254	0.0093
L_1-SCN	9.8766	0.3888	7.2653	0.2773	0.9467	0.0042
DC-SCN	9.5123	0.5362	7.1025	0.3282	0.9504	0.0056
SR-SCN	**9.2507**	**0.2910**	**6.9851**	**0.2314**	**0.9532**	**0.0030**

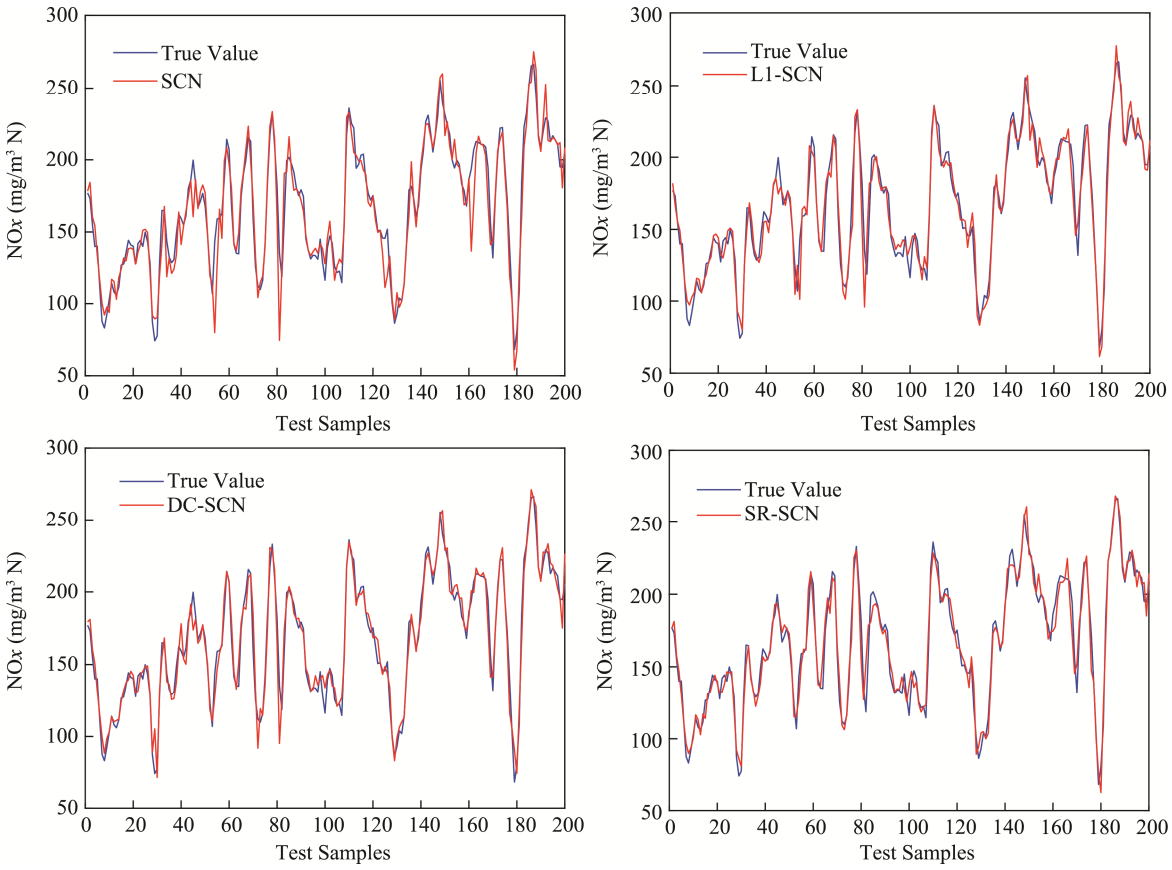

Fig.5 Comparison of the fitting results of different prediction models

4 Conclusion

To achieve accurate prediction of NO*x* emission concentration in the MSWI process, this paper proposes a modeling method based on a sparse regularization stochastic configuration network (SR-SCN) in response to the SCN algorithm modeling's overfitting and poor generalizability problems. The method applies DropConnect regularization to the input weights to break

the co-adaptation between nodes, and combines L_1 regularization to constrain the output weights and eliminate the redundant nodes, which implements the sparsification of the SCN prediction model. Experiments with standard datasets and actual data from MSWI plant proved that SR-SCN can effectively resolve the overfitting problems caused by complex network structures. Compared with the SCN, DC-SCN, and L_1-SCN methods, the SR-SCN method has greater accuracy and generalizability for NO*x* concentration prediction and effectively simplifies the model structure, which is valuable for the research and application of pollutant prediction in the MSWI process. Since SR-SCN uses the ADMM algorithm to iteratively determine the output weights, which increases the computational complexity of the model, designing a lightweight training algorithm based on the proposed method to optimize the computational efficiency and resource consumption will be the focus of future research.

Author Contributions:

Contributions：YAN Aijun: Project administration , Supervision, Writing - review & editing. CAO Shenci: Writing - original draft and Writing - review & editing.

Funding Information:

This work is supported by the National Natural Science Foundation of China (62373017, 62073006) and the Beijing Natural Science Foundation of China (4212032).

Data Availability:

The data that support the results of this study are available upon reasonable request by contacting the corresponding author.

Conflict of Interest:

The authors declare no competing interests.

Dates:

Received 2 June 2024; Accepted 7 July 2024; Published online 30 September 2024

References

[1] Li Y, Zhao X G, Li Y B, Li X (2015). Waste incineration industry and development policies in China. *Waste Management*, 46(12), pp. 234-241.

[2] Kinga S, Jacek S M, Stanislaw L (2010). Trends in NO abatement: A review. *Science of the Total Environment*, 408(19), pp. 3976-3989.

[3] Si M X, Du K (2020). Development of a predictive emissions model using a gradient boosting machine learning method. *Environmental Technology & Innovation*, 20(11), pp. 101028.

[4] Qiao J F, Zhou J L, Meng X (2023). A multitask learning model for the prediction of NOx emissions in municipal solid waste incineration processes. *IEEE Transactions on Instrumentation and Measurement*, 72(4), pp. 1-14.

[5] Wang F, Ma S X, Wang H, Li Y D, Zhang J J (2018). Prediction of NOx emission for coal-fired boilers based on deep belief network. *Control Engineering Practice*, 80(11), pp. 26-35.

[6] Wang G Y, Liu S Y, Shuai S J, Wang Z M (2020). NOx emissions prediction based on mutual information and back propagation neural network using correlation quantitative analysis. *Energy*, 198(5), pp. 117286.

[7] Liu G F, Zhang Y, Shen D K, Yuan B, Li R, Sun Y (2021). Anticipatory NH3 injection control for SCR system based on the prediction of the inlet NOx concentration. *Journal of the Energy Institute*, 94(2), pp. 167-175.

[8] Alhamdoosh M, Wang D (2014). Fast decorrelated neural network ensembles with random weights. *Information Sciences*, 264(4), pp. 104-117.

[9] Pao Y H, Takefuji Y (1992). Functional-link net computing: theory, system architecture, and functionalities. *Computer*, 25(5), pp. 76-79.

[10] Gorban A N, Tyukin I Y, Prokhorov D V, Sofeikov K I (2016). Approximation with random bases: Pro et Contra. *Information Sciences*, 364(10), pp. 129-145.

[11] Li M, Wang D H (2017). Insights into randomized algorithms for neural networks: Practical issues and common pitfalls. *Information Sciences*, 382(3), pp. 170-178.

[12] Wang D H, Li M (2017). Stochastic configuration networks: fundamentals and algorithms. *IEEE Transactions on Cybernetics*, 47(10), pp. 3346-3479.

[13] Wang S Q, Lian G G, Cheng C, Chen H T (2024). A novel method of rolling bearings fault diagnosis based on singular spectrum decomposition and optimized stochastic configuration network. *Neurocomputing*, 574(3), pp. 127278.

[14] Yan A J, Wang R R, Guo J C, Tang J (2024). A knowledge transfer online stochastic configuration network-based prediction model for furnace temperature in a municipal solid waste incineration process. *Expert Systems with Applications*, 243, pp. 122733.

[15] Yan A J, Guo J C, Wang D H (2022). Heterogeneous feature ensemble modeling with stochastic configuration networks for predicting furnace temperature of a municipal solid waste incineration process. *Neural Computing and Applications*, 34(18), pp. 15807-15819.

[16] Li K, Zhang L M, Qiao J F (2024). Multi-task stochastic configuration network with autonomous linking and its application in wastewater treatment processes. *Information*

Sciences, 662(3), pp. 120195.

[17] Yan A J, Guo J C, Wang D H (2022). Robust stochastic configuration networks for industrial data modelling with student's-t mixture distribution. *Information Sciences*, 607(8), pp. 493-505.

[18] Qian X S, Huang H, Chen X P, Huang T W (2017). Efficient construction of sparse radial basis function neural networks using L_1-regularization. *Neural Networks*, 94(10), pp. 239-254.

[19] 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(1), pp. 1929-1958.

[20] Wan L, Zeiler M, Zhang S, Lecun Y, Fergus R (2013). Regularization of neural networks using DropConnect. *International Conference on Machine Learning*, 28(6), pp. 1058-1066.

[21] Hssayni E H, Joudar N E, Ettaouil M (2022). An adaptive Drop method for deep neural networks regularization: Estimation of DropConnect hyperparameter using generalization gap. *Knowledge-Based Systems*, 253(10), pp. 109567.

[22] Wang Q J, Dai W, Lu Q, Fu X R, Ma X P (2022). A sparse learning method for SCN soft measurement model. *Control and Decision*, 37(12), pp. 3171-3182.

[23] Lu J, Ding J L, Liu C X, Chai T Y (2022). Hierarchical-Bayesian-based sparse stochastic configuration networks for construction of prediction intervals. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8), pp. 3560-3571.

[24] Ho T K (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), pp. 832-844.

Peng H C, Long F H (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(8), pp. 1226-1238.