China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 12, No. 4, December 2025
https://doi.org/10.15878/j.instr.202500269

Article

A Stochastic Configuration Network Modeling Method Based on Improved Hidden Layer Output Matrix and Supervisory Mechanism

Aijun Yan[1,2], Jing Wang[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

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

Abstract: To improve the generalization performance and prediction accuracy of the stochastic configuration network (SCN) model, a novel SCN modeling method is proposed. First, the first- and second-order directional derivatives of the hidden layer output matrix are calculated. The key factors extracted from the directional derivatives are linearly added to the original hidden layer output matrix to formulate a new hidden layer output matrix. Second, a spatial angle adaptive supervisory mechanism is established to improve the quality of the parameter configuration of the hidden layer nodes. The experimental results show that the proposed method improves the generalization performance and prediction accuracy. This work is a beneficial exploration of the standard SCN algorithm.

Keywords: stochastic configuration network; directional derivative; angle adaptive; municipal solid waste incineration; flue gas oxygen content

Citation: Aijun Yan, Jing Wang."A Stochastic Configuration Network Modeling Method Based on Improved Hidden Layer Output Matrix and Supervisory Mechanism." Instrumentation 12, no.4 (December 2025). https://doi.org/10.15878/j.instr.202500269

1 Introduction

With increased computing power and data storage capabilities, data-driven modeling methods have been widely applied[1]. The artificial neural network (ANN) is a fast and efficient data-driven modeling approach that has been successfully applied in the field of industrial process prediction modeling[2,3]. However, back propagation (BP) neural networks and radial basis function (RBF) neural networks based on the gradient descent algorithm for parameter learning have shortcomings, such as easily converging to local optima and slow convergence speeds. These issues can limit the modeling speed and accuracy of ANN models. Therefore, further research is needed on how to construct a neural network that is both faster and more accurate[4].

The random learning technique in ANNs involves randomly assigning input weights and biases. It has the advantages of simple implementation and fast learning speed, and it can be applied to neural network modeling while demonstrating good generalization performance[5]. In[6], a random vector functional link (RVFL) network based on the idea of random learning has been proposed, and it has been applied in various fields[7,8]. However, the universal approximation ability of the RVFL network has not been rigorously theoretically proven; that is, owing to the random configuration of hidden layer parameters within a fixed range, its universal approximation ability needs to be improved, and the network structure is difficult to determine[9]. To solve these problems, Wang and Li proposed a stochastic configuration network (SCN), for which the hidden layer parameters are constrained by adding a supervisory mechanism to ensure the universal approximation ability[10]. This method has been applied in industrial process parameter modeling fields such as air flow monitoring[11], hematite particle size estimation[12], and furnace temperature prediction[13].

In the SCN modeling process, two factors affect the generalization performance and accuracy of the model. First, the stochastic configuration network constructed with a single kernel function has a weak mapping capability on the basis of the hidden layer output matrix, which may affect the generalization performance of the

model. In[14], L_1 regularization was introduced during the incremental construction of the stochastic configuration network, and a new inequality constraint was established to enhance the quality of the hidden layer nodes, avoid overfitting caused by complex structures, and improve the sparsity and generalization ability of the model. In[15], an improved SCN algorithm based on the L_2 regularization technique was introduced, and the accuracy and generalization ability of the original model were enhanced. However, the hidden layer output matrices used in the above studies were all obtained through single kernel function mappings, which may have a certain impact on model generalization performance. Second, the key influential factor is the constraint degree of the supervisory mechanism. The constraint degree of the SCN's supervisory mechanism is set by adaptively searching for the appropriate parameter r based on the variation in the sample data. By adjusting the value of r, nodes that can maximize the reduction in the residual are selected from the candidate pool. An inappropriate value of r may lead to a high node rejection rate in the SCN, thereby affecting the modeling accuracy. In[16], an improved supervisory mechanism was introduced by reconstructing the parameter r, which affects the constraint degree of the supervisory mechanism; consequently, the parameter r was adaptively varied according to the number of hidden layer nodes, thereby improving the modeling efficiency of the SCN. In[17], an adaptive change supervisory mechanism was adopted to allocate hidden layer parameters, which improved the efficiency of model parameter configuration. By reconstructing the key parameters that affect the constraint degree of the supervisory mechanism, the modeling performance of the SCN can be improved.

In summary, a stochastic configuration network modeling method based on an improved hidden layer output matrix and supervisory mechanism (IHS-SCN) is proposed to enhance the generalization performance and prediction accuracy of SCN models. The main contributions of the paper are as follows: 1) By calculating the directional derivative of the original hidden layer output matrix, the key parts of the directional derivative that affect the model's generalization performance are linearly added to the original hidden layer output matrix to construct a new hidden layer output matrix, thereby enhancing the model's generalization performance. 2) Based on the spatial geometric angle between the residual and the activation function, a new angle adaptive supervisory mechanism is constructed, which can adaptively change the constraint degree of the supervisory mechanism to improve the accuracy of the model. 3) The above two improvements are applied to the traditional SCN model, and the effectiveness of the proposed method is verified through experiments.

The rest of the paper is organized as follows: Section 2 introduces the principle of the SCN and the analysis of related problems; Section 3 describes the modeling process of the IHS-SCN model; Section 4 provides the experimental validation and the analysis of results; and Section 5 concludes the paper.

2 Related Work

This section introduces the algorithmic principles of SCN and discusses the problems existing in its modeling process.

2.1 SCN Algorithm

The SCN is an incrementally constructed single hidden layer neural network that starts with one node. Then, the number of nodes is gradually increased through the constraints of the supervisory mechanism, and the least squares method is used to calculate the output weights until the preset accuracy is achieved and the network is completed. The construction process is as follows.

For the objective function f: $\mathbb{R}^d \to \mathbb{R}^m$, given a dataset D= $\{X, Y\}$, where the input variables are X=$\{x_1, x_2, \cdots, x_N\}$, N is the total number of samples, x_i=$\{x_{i,1}, \cdots, x_{i,d}\}^{\mathrm{T}}$ $\in \mathbb{R}^d$, and the output variables are Y= $\{y_1, y_2, \cdots, y_N\}$, y_i= $\{y_{i,1}, \cdots, y_{i,m}\}^{\mathrm{T}} \in \mathbb{R}^m$, assuming that L-1 hidden layer nodes are established, the output of the SCN can be expressed as:

$$f_{L-1}(X) = \sum_{j=1}^{L-1} \beta_j^* g_j(w_j^{\mathrm{T}} X + b_j) \tag{1}$$

where β_j^* represents the output weight of the jth hidden layer node, $g_j(\cdot)$ is the activation function of the jth hidden layer node, w_j and b_j are the input weight and bias of the jth node, respectively, and f_{L-1} is the current model output.

The current network residual is as follows:

$$e_{L-1}^* = Y - f_{L-1} = [e_{L-1}^*(x_1), ..., e_{L-1}^*(x_N)]^{\mathrm{T}} \tag{2}$$

During each iteration, the SCN adds a new hidden layer node obtained by maximizing the supervisory mechanism to reduce the network residuals. If the current network residual does not reach the preset accuracy or the maximum number of hidden layer nodes $L_{\max}$ is not reached, the hidden layer parameters are determined according to the supervisory mechanism constraint as expressed in (4):

$$h_L = [g_L(w_L^{\mathrm{T}} x_1 + b_L), ..., g_L(w_L^{\mathrm{T}} x_N + b_L)]^{\mathrm{T}} \in R^N \tag{3}$$

$$\xi_{L,q} = \frac{\left\langle e_{L-1,q}^*, h_L \right\rangle^2}{h_L^T h_L} - (1 - r - \mu_L) \left\| e_{L-1,q}^* \right\|^2 \tag{4}$$

where q=1, 2, $\cdots$, m; h_L represents the output of hidden layer node L; w_L and b_L are the candidate parameters for hidden layer node L; $r \in (0, 1)$; $\{\mu_L\}$ denotes a sequence of nonnegative real numbers, with $\lim_{L\to\infty} \mu_L = 0$ and $\mu_L \le (1-r)$; and the candidate node parameters that satisfy the maximum value constraint $\xi_L = \sum_{q=1}^{m} \xi_{L,q} \ge 0$ are used as the Lth node parameters.

The output weights of the hidden layer nodes are determined according to the following equation until the network meets the preset accuracy:

$$\beta^* = \arg\min_{\beta^*} \left\| H_L\beta - Y \right\|^2 = H_L^+ Y \tag{5}$$

where $H_L = [h_1, h_2, \cdots, h_L]$ represents the hidden layer output matrix and $(\cdot)^+$ denotes the pseudoinverse operation.

The network output result f is as follows:

$$f = H\beta^* \tag{6}$$

2.2 Problem Analysis

Compared with BP and RBF networks, the SCN has advantages such as simple implementation and high modeling efficiency[11-17]. However, there are several issues that affect its generalization ability and modeling precision.

1) The use of a single kernel function in the hidden layer can affect its generalization performance. To ensure the nonlinear mapping capability of the model, a single kernel function in the SCN hidden layer is used to map the input feature space to a high-dimensional feature space. However, for different modeling tasks, the maximum number of hidden layer nodes in the model may not meet the expected tolerance. In this case, it is necessary to replace the kernel function with a stronger mapping function or increase the number of hidden layer nodes to improve the expression ability of the model. However, increasing the number of hidden layer nodes may increase the complexity of the model and easily lead to overfitting. Therefore, further research is needed to enhance the generalization ability of the model from the perspective of enhancing the hidden layer mapping capability.

2) The quality of the hidden layer parameter configuration is not high. Notably, this quality is influenced by the constraint degree of the supervisory mechanism, which is determined by the variable parameter r. The variable parameter r of the SCN is selected within a fixed interval, and this selection method may not have sufficient constraint ability on its hidden layer nodes, resulting in low quality of newly added hidden layer nodes and affecting the modeling accuracy. Therefore, further research is needed to select an appropriate constraint degree to improve the model accuracy.

3 Method

To address the abovementioned problems of single hidden layer mapping and low-quality parameter configuration in the traditional SCN, a stochastic configuration network modeling method based on an improved hidden layer output matrix and supervisory mechanism (IHS-SCN) is proposed. This method enhances the mapping ability of the hidden layer by reconstructing the hidden layer output matrix; in addition, the supervisory mechanism is improved on the basis of the spatial angle variation between residuals and activation functions to enhance the quality of the hidden layer parameter configuration.

3.1 Improving the Hidden Layer Output Matrix

SCN is a single hidden layer feedforward neural network. In different modeling tasks, a phenomenon by which the maximum number of nodes does not meet the expected tolerance may occur because of the limitations of single hidden layer mapping. Therefore, it is necessary to enhance the mapping ability of the hidden layer output matrix and improve the generalization ability of the model[18]. Since calculating the derivative of the hidden layer output matrix can yield valuable information about the mapping and the directional derivative does not expand the dimensionality of the original matrix, calculating the directional derivative of the hidden layer output matrix can yield more information about the key factors affecting the model's generalization performance. On the basis of the idea of second-order Taylor series expansion of the hidden layer output matrix in[19], the first-order and second-order directional derivatives of the hidden layer output matrix are calculated, and then, the key factors affecting the model's generalization ability in the first-order and second-order directional derivatives are considered. Therefore, these key factors are linearly added to the original hidden layer output matrix to construct a new hidden layer output matrix, which improves the model's generalization ability.

The hidden layer output matrix H is assumed to be a differentiable matrix with a value function, i. e., $H(X) \in \mathbb{R}^{N\times d} \to \mathbb{R}^{N\times L}$, as shown in (7). This matrix is differentiable to a continuous first-order degree, as shown in (8)-(9):

$$H(X) = \begin{bmatrix} g(w_1^{\mathrm{T}}x_1 + b_1) & \cdots & g(w_L^{\mathrm{T}}x_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(w_1^{\mathrm{T}}x_N + b_1) & \cdots & g(w_L^{\mathrm{T}}x_N + b_L) \end{bmatrix} \tag{7}$$

$$\nabla H(X) = \begin{bmatrix} \nabla H_{1,1}(X) & \cdots & \nabla H_{1,L}(X) \\ \vdots & \cdots & \vdots \\ \nabla H_{N,1}(X) & \cdots & \nabla H_{N,L}(X) \end{bmatrix} \tag{8}$$

$$\nabla H_{i,j}(X) = \begin{bmatrix} \dfrac{\partial g(w_j^{\mathrm{T}}x_i + b_j)}{\partial x_{1,1}} & \cdots & \dfrac{\partial g(w_j^{\mathrm{T}}x_i + b_j)}{\partial x_{1,d}} \\ \vdots & \cdots & \vdots \\ \dfrac{\partial g(w_j^{\mathrm{T}}x_i + b_j)}{\partial x_{N,1}} & \cdots & \dfrac{\partial g(w_j^{\mathrm{T}}x_i + b_j)}{\partial x_{N,d}} \end{bmatrix} \tag{9}$$

where $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, L$.

Assuming that there is sufficiently small random noise in the selected sample matrix X, i. e., $\tilde{X} = X + \lambda Z$, where $Z \in \mathbb{R}^{N\times d}$ is a random matrix and λ is an arbitrary constant sufficiently small, the expansion of the second-order Taylor series with respect to $\tilde{X}$ is

$$H(X + \lambda Z) = H(X) + \lambda \overrightarrow{d}^{Z} H(X) + \frac{\lambda^2}{2!} \overrightarrow{d}^{Z} H^2(X) + o(\lambda^3) \tag{10}$$

The first-order and second-order directional derivatives of the hidden layer output matrix $H(X)$ in the given $Z \in \mathbb{R}^{N\times d}$ direction can be expressed as:

$$\overrightarrow{d}^{Z} H(X)= \begin{bmatrix} tr(\nabla H_{1,1}(X)^{\mathrm{T}}Z) & \cdots & tr(\nabla H_{1,L}(X)^{\mathrm{T}}Z) \\ \vdots & \cdots & \vdots \\ tr(\nabla H_{N,1}(X)^{\mathrm{T}}Z) & \cdots & tr(\nabla H_{N,L}(X)^{\mathrm{T}}Z) \end{bmatrix} \tag{11}$$

$$\overrightarrow{d}^{Z} H^{2}(X)= \begin{bmatrix} tr(\nabla tr(\nabla H_{1,1}(X)^{\mathrm{T}}Z)^{\mathrm{T}}Z) & .. & tr(\nabla tr(\nabla H_{1,L}(X)^{\mathrm{T}}Z)^{\mathrm{T}}Z) \\ \vdots & .. & \vdots \\ tr(\nabla tr(\nabla H_{N,1}(X)^{\mathrm{T}}Z)^{\mathrm{T}}Z) & .. & tr(\nabla tr(\nabla H_{N,L}(X)^{\mathrm{T}}Z)^{\mathrm{T}}Z) \end{bmatrix} \tag{12}$$

In this paper, the sigmoid function is chosen as the activation function, i. e., $g(t) = 1/(1+e^{-t})$, with its first derivative $g'(t)$ and second derivative $g''(t)$. Therefore, the first-order directional derivative of the hidden layer output matrix with input weights and biases can be expressed as shown in (13) - (14), and the second-order directional derivative is shown in (15)-(16):

$$\begin{aligned} tr(\nabla H_{i,j}(X)Z) &= \sum_{i'=1}^{N}\sum_{k'=1}^{d}\frac{\partial g(w_j^{\mathrm{T}}x_i+b_j)}{\partial x_{i',k'}}Z_{i',k'} \\ &= g(w_j^{\mathrm{T}}x_i+b_j)(1-g(w_j^{\mathrm{T}}x_i+b_j))\sum_{k=1}^{d}w_{j,k}Z_{i,k} \end{aligned} \tag{13}$$

$$\begin{aligned} \overrightarrow{d}^{Z} H(X) &= \sum_{i=1}^{N}\sum_{j=1}^{L} tr(\nabla H_{i,j}(X)^{\mathrm{T}}Z) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{L} g_{ij}(1-g_{ij})\sum_{k=1}^{d}w_{j,k}Z_{i,k} \\ &= H\circ(O-H)\circ\ddot{W}\cdot Z \end{aligned} \tag{14}$$

$$\begin{aligned} &tr(\nabla tr(\nabla H_{i,j}(X)^{\mathrm{T}}Z)^{\mathrm{T}}Z) \\ &= \sum_{i''=1}^{N}\sum_{k''=1}^{d}\frac{\partial tr(\nabla H_{i,j}(X)^{\mathrm{T}}Z)}{\partial x_{i'',k''}}Z_{i'',k''} \\ &= \sum_{i'=1}^{N}\sum_{k'=1}^{d}\sum_{i''=1}^{N}\sum_{k''=1}^{d}\frac{\partial^2 g(w_j^{\mathrm{T}}x_i+b_j)}{\partial x_{i',k'}\partial x_{i'',k''}}Z_{i',k'}Z_{i'',k''} \\ &= g_{ij}(1-g_{ij})(1-2g_{ij})\sum_{k'=1}^{d}\sum_{k''=1}^{d}w_{j,k'}w_{j,k''}Z_{i,k'}Z_{i,k''} \end{aligned} \tag{15}$$

$$\begin{aligned} \overrightarrow{d}^{Z} H^{2}(X) &= \sum_{i=1}^{N}\sum_{j=1}^{L} tr(\nabla tr(\nabla H_{i,j}(X)^{\mathrm{T}}Z)^{\mathrm{T}}Z) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{L} g_{ij}(1-g_{ij})(1-2g_{ij})\sum_{k'=1}^{d}\sum_{k''=1}^{d}w_{j,k'}w_{j,k''}Z_{i,k'}Z_{i,k''} \\ &= H\circ(O-H)\circ(O-2H)\circ\ddot{W}\circ\ddot{W}\cdot Z\cdot Z \end{aligned} \tag{16}$$

where $O \in \mathbb{R}^{N\times L}$ is the all-1 matrix, $\ddot{W} \in \mathbb{R}^{N\times L}$ is the row vector $(w_1, w_2, \ldots w_L)$ replicated N times, and $\circ$ represents the Hadamard product.

The second-order Taylor series expansion is sufficient to identify some key factors when constructing the model. As indicated by (10), (14), and (16), the main factors affecting the hidden layer output matrix are H, $H\circ(O-H)\circ\ddot{W}$ and $H\circ(O-H)\circ(O-2H)\circ\ddot{W}\circ\ddot{W}$. Therefore, by linearly adding these key influencing factors, a new hidden layer output matrix can be obtained:

$$\begin{aligned} H^{*} &= H + H\circ(O-H)\circ\ddot{W} \\ &+ H\circ(O-H)\circ(O-2H)\circ\ddot{W}\circ\ddot{W} \end{aligned} \tag{17}$$

Remark 1:

While the incorporation of higher-order directional derivatives increases model complexity to some extent, it reveals more key factors that influence generalization performance, thereby guiding the design of optimization strategies for improving model performance. To reduce the computational burden, only second-order directional derivatives are adopted in this paper, achieving a trade-off between predictive accuracy and computational cost.

3.2 Improving the Supervisory Mechanism

The SCN constrains the hidden layer parameters through the supervised mechanism, ensuring its universal approximation capability. The supervisory mechanism of the SCN controls the tightness of constraints by changing the parameter r value. As the number of hidden layer nodes increases, the value of parameter r should be as close to 1 as possible. Therefore, the selectable range for the r value becomes smaller, and the constraint degree of the supervisory mechanism is limited, resulting in low quality of hidden layer nodes and affecting the accuracy of SCN modeling. In[18], the residual e_L and activation function g_L were analyzed from the perspective of spatial geometry, and it was found that maximizing the spatial angle θ_L enhances model convergence. Starting from the perspective of enhancing the constraint degree of the supervisory mechanism, the concept of spatial perspective is introduced into the supervisory mechanism and improves it, as shown in (19); therefore, the constraint degree of the supervisory mechanism changes with the variation of spatial angle θ_L.

Theorem 1[18]**:** Suppose that span(Γ) is dense in L_2 space and $\forall\ g \in \Gamma$, $0 < \|g\| < b_g$ for some $b_g \in \mathbb{R}^{+}$. Intermediate values $\tilde{\beta}_L = \langle e_{L-1}^{*}, g_L\rangle / \|g_L\|^2$ are defined, and $\tilde{e}_L = e_{L-1}^{*} - \tilde{\beta}_L g_L$, from the perspective of the function space:

$$\begin{aligned} \langle \tilde{e}_L, g_L\rangle &= \langle \tilde{e}_{L-1} - \beta_L g_L, g_L\rangle \\ &= \langle \tilde{e}_{L-1}, g_L\rangle - \beta_L\langle g_L, g_L\rangle \\ &= \langle \tilde{e}_{L-1}, g_L\rangle - \langle \tilde{e}_{L-1}, g_L\rangle \\ &= 0 \end{aligned} \tag{18}$$

Then, we can obtain $\tilde{e}_L \perp g_L$.

Since $\tilde{e}_L \perp g_L$ and $\tilde{e}_L = e_{L-1}^{*} - \tilde{\beta}_L g_L$, from a spatial perspective, the relationship among $\tilde{e}_L$, e_{L-1}^{*} and g_L is shown in Fig. 1 The angle $\angle(e_{L-1}^{*}, g_L) \in (0, \pi/2)$ between e_{L-1}^{*} and g_L is represented by θ_L, which changes with increasing number of nodes. By using the change of θ_L to improve the supervisory mechanism, it is interpretable in space.

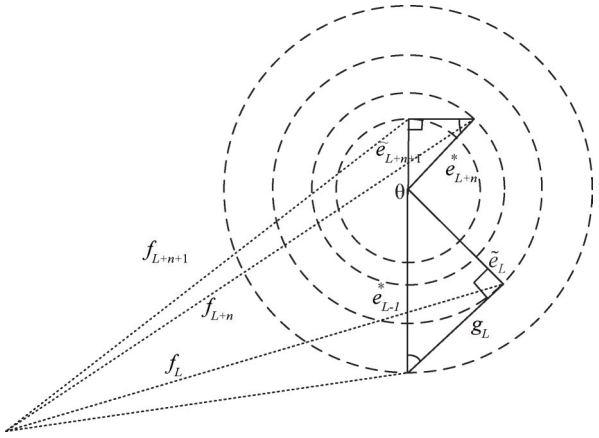


Fig.1 Schematic diagram of the relationship between the residual and the activation function

Theorem 2: Suppose that span(Γ) is dense in L_2 space and $\forall g\in\Gamma$, $0<\|g\|<b_g$ for some $b_g\in\mathbb{R}^+$. For L=1, 2, ⋯, let$(1-\sin\theta_L)^\alpha<\cos^3\theta_L$, where $\alpha\in\mathbb{R}^+$. If the newly added hidden layer nodes satisfy the inequality constraint shown in (19) and the output weights are calculated as shown in (5), then $\lim_{L\to+\infty}\|f-f_L\|=0$.

$$\xi_L=\frac{\left\langle e^*_{L-1},g_L\right\rangle^2}{\|g_L\|^2}-(1-\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L})\|e^*_{L-1}\|^2>0 \quad (19)$$

Proof:

It is easy to show that $\|e^*_L\|^2\le\|\tilde{e}_L\|^2=\|e^*_{L-1}-\tilde{\beta}_L g_L\|^2\le\|\tilde{e}^*_{L-1}\|^2\le\|\tilde{e}_{L-1}\|^2$, so $\|e^*_L\|^2$ is monotonically decreasing and convergent. Suppose that $\lim_{L\to+\infty}\|e^*_L\|^2$; then, we have

$$\begin{aligned}
&\|e^*_L\|^2-\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L}\|e^*_{L-1}\|^2\\
&\le\|\tilde{e}_L\|^2-\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L}\|e^*_{L-1}\|^2\\
&=\left\langle e^*_{L-1}-\tilde{\beta}_L g_L,e^*_{L-1}-\tilde{\beta}_L g_L\right\rangle-\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L}\|e^*_{L-1}\|^2\\
&=(1-\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L})\|e^*_{L-1}\|^2-2\left\langle e^*_{L-1},\tilde{\beta}_L g_L\right\rangle+\left\langle\tilde{\beta}_L g_L,\tilde{\beta}_L g_L\right\rangle\\
&=(1-\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L})\|e^*_{L-1}\|^2-\frac{\left\langle e^*_{L-1},g_L\right\rangle^2}{\|g_L\|^2}
\end{aligned} \quad (20)$$

Due to the constraints imposed by the supervisory mechanism in Eq. (19) and$(1-\sin\theta_L)^\alpha<\cos^3\theta_L$, we obtain $\|e^*_L\|^2\le\frac{(1-\sin\theta_L)^\alpha}{\cos\theta_L}\|e^*_{L-1}\|^2<\cos^2\theta_L\|e^*_{L-1}\|^2$. Given that $\tilde{e}_L\perp g_L$ and $\tilde{e}_L=e^*_{L-1}-\tilde{\beta}_L g_L$, since $\tilde{\beta}_L$ is a numerical value, it can be deduced that $\tilde{e}_L\perp(e^*_{L-1}-\tilde{e}_L)$. Therefore,

$$\begin{aligned}
\|e^*_{L-1}-\tilde{e}_L\|^2&=\|e^*_L\|^2+\|\tilde{e}_L\|^2-\left\langle e^*_{L-1},\tilde{e}_L\right\rangle\\
&=\|e^*_{L-1}\|^2+\|\tilde{e}_L\|^2-\left(\left\langle e^*_{L-1},\tilde{e}_L\right\rangle-\left\langle e^*_{L-1}-\tilde{e}_L,\tilde{e}_L\right\rangle\right)\\
&=\|e^*_{L-1}\|^2-\|\tilde{e}_L\|^2
\end{aligned} \quad (21)$$

Thus, $\lim_{L\to+\infty}\|\tilde{\beta}_L g_L\|^2=\lim_{L\to+\infty}(\|e^*_{L-1},\tilde{e}_L\|)=p^2-p^2=0$. From the spatial perspective, it can be concluded that when $\lim_{L\to+\infty}\|\tilde{\beta}_L g_L\|=0$, $\lim_{L\to+\infty}\theta_L=\pi/2$, $\lim_{L\to+\infty,\theta_L=\pi/2}\cos^2\theta_L\|e^*_{L-1}\|=0$; hence, $\lim_{L\to+\infty}\|e^*_L\|=0$.

3.3 Pseudocode of the IHS-SCN Algorithm

On the basis of the above, the pseudocode of the IHS-SCN algorithm is as follows.

Input: Dataset $D=\{X, Y\}$, the maximum number of hidden layer nodes $L_{\max}$, the maximum number of configurations $T_{\max}$, the tolerance error ε , and the hidden layer parameter configuration range $\Upsilon=[\lambda_{\min}:\Delta\lambda:\lambda_{\max}]$.

1. Initialize: $e_0=[y_1, y_2, \cdots, y_N]^T$;
2. While $L\leqslant L_{\max}$ and $\|e_0\|>\varepsilon$
3. Assign input weights w_j and biases b_j from Υ;
4. Calculate ξ_L based on Eq. (17) and Eq. (19);
5. Find w_j and b_j that maximizeξ_L;
6. Calculate H^*_L by Eq.(7), calculateβ^* by Eq.(5) ,
7. Renew $e_L=Y-H^*_L\beta^*$, $L=L+1$;
8. End While

Output: IHS-SCN model

4 Experiments and Analysis of Results

This paper introduces directional derivatives into the calculation of the hidden layer output matrix, and the spatial angle changes between the residuals and activation functions are considered when adjusting the constraint degree of the supervisory mechanism to improve the generalization performance and prediction accuracy of the SCN model. To verify the effectiveness of the proposed method, ablation experiments and comparative experiments were designed. The ablation experiments involved comparing the IHS-SCN with the IH-SCN (SCN based on an improved hidden layer output matrix) and IS-SCN (SCN based on an improved supervisory mechanism). Additionally, in comparative experiments, the IHS-SCN was compared with RVFL, SCN, PSCN[14], SGA-SIM-I[18], and CIRWN[21]. For convenience of description, SCN and its variant models are referred to as SCN-type models. The experimental data consisted of four standard datasets from the KEEL[1] database and one historical dataset for municipal solid waste incineration, with specific details provided in Table 1. The datasets were divided into 60% for training, 20% for validation, and 20% for testing, with all the data normalized to the interval [0, 1].

[1]KEEL: http://www.keel.es/

Table 1 Dataset information

Dataset	Samples	Inputs	Outputs
AutoMPG8	392	7	1
Stock	950	9	1
Concrete	1030	8	1
ANACALT	4052	7	1
MSWI	1000	31	1

4.1 Experimental Preparation

The root mean square error (RMSE), mean absolute error (MAE), and determination coefficient (R^2) are selected as the evaluation metrics for the experimental results, and the corresponding formulas are as follows:

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{22}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|y_i-\hat{y}_i\right| \tag{23}$$

$$\mathrm{R}^2=1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2} \tag{24}$$

where y_i and $\hat{y}_i$ are the true and predicted values for the ith sample, $\bar{y}_i$ is the mean of the true values, and N is the number of samples.

The parameter settings for each model are shown in Table 2. The r values for the supervisory mechanisms in the SCN and IH-SCN are [0.9, 0.99, ⋯, 0.999999], and the activation functions for each model are sigmoid functions with a tolerance error of 0.01. All the hyperparameters are obtained from the validation set results.

Table 2 Parameter settings for each model

Dataset	RVFL (L_{max}, T_{max}, λ)	SCN-type models CIRWN SGA-SIM-I (L_{max}, T_{max})	SCN-type models CIRWN SGA-SIM-I (Υ)	IHS-SCN (α)
AutoMPG8	25, 1, 1	25, 100	[0.5, 1, 5, 10, 30, 50, 100, 150, 200, 250]	1.4
Stock	50, 1, 1	50, 100		1.35
Concrete	50, 1, 1	50, 100		1.4
ANACALT	50, 1, 1	50, 100		1.3
MSWI	100, 1, 1	100, 100		1.4

4.2 Ablation Experiment

Fig. 2 presents the ablation experiment results, which are the testing set evaluation metrics for SCN, IH-SCN, IS-SCN, and IHS-SCN on four standard datasets and are used to verify the effectiveness of different improvement strategies. As shown in Fig. 2, the overall performance of the IHS-SCN is superior to that of the other three models. For the four datasets, the IH-SCN performs better than the SCN does, indicating the effectiveness of using directional derivatives to improve the hidden layer output matrix strategy. For the AutoMPG8, Concrete, and ANACALT datasets, the performance of the IS-SCN is superior to that of the SCN, indicating the effectiveness of the strategy of improving the supervisory mechanism with an adaptive angle strategy. For the Stock dataset, the IS-SCN is slightly inferior to the SCN, indicating that improvements to the SCN supervisory mechanism should be explored in conjunction with the characteristics of the dataset itself.

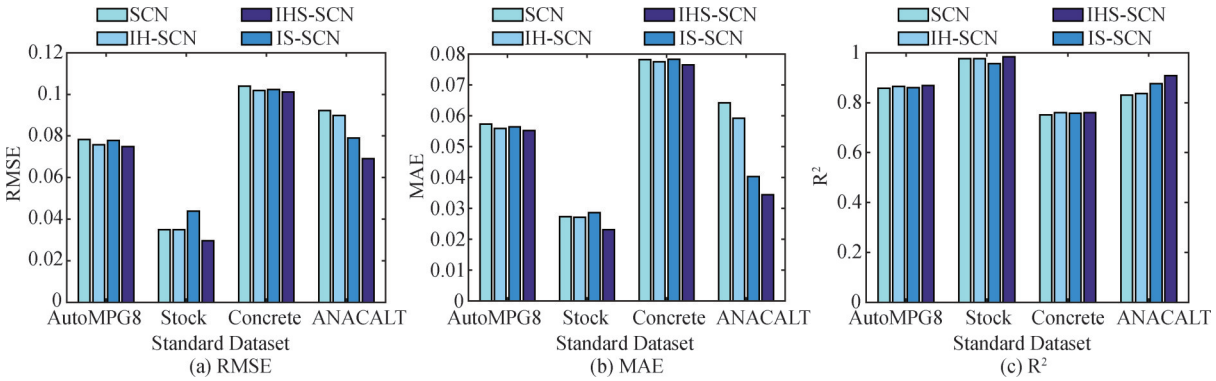


Fig.2 Results of the ablation experiment.

4.3 Comparison Experiment of Standard Datasets

In this section, comparative experiments were conducted on the basis of four standard datasets: AutoMPG8, Stock, Concrete, and ANACALT. Each model was run independently 30 times, and the mean values of the evaluation metrics were selected for

comparison, as shown in Table 3. Table 3 shows that the SCN outperforms the RVFL in both the training and testing sets on four datasets, indicating that using the supervised mechanism to constrain the hidden layer parameters can effectively improve the performance of the model. Compared with those of the SCN, the evaluation metrics of the PSCN are superior to those of the SCN for the Stock, Concrete, and ANACALT datasets and comparable to those of the SCN for the AutoMPG8 dataset, indicating that using sparse constraints can improve the generalization performance of the model. The evaluation metrics of SGA-SIM-I are superior to those of the SCN for the AutoMPG8, Stock, and ANACALT datasets and comparable to those of the SCN for the Concrete dataset, indicating that the use of a spatial angle maximization improvement strategy can effectively enhance the performance of the model. The evaluation metrics of the CIRWN are slightly better than those of the SCN for the AutoMPG8 and Stock datasets and are comparable to those of SCN for the Concrete and ANACALT datasets, indicating that using a strict supervisory mechanism for constraints can effectively improve the accuracy of the model.

Table 3 Comparison of the evaluation metrics for each model

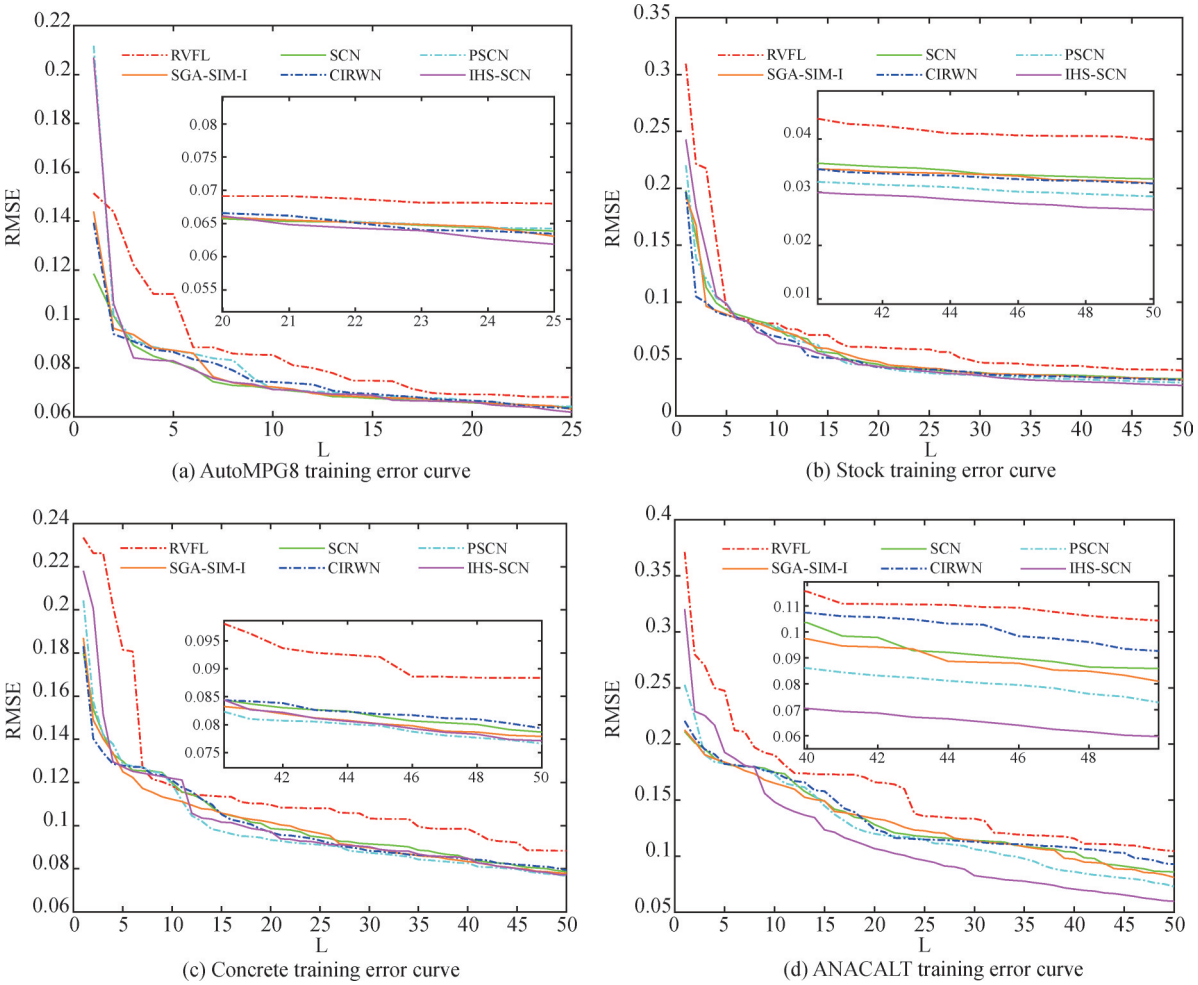
Dataset	Metric		RVFL	SCN	PSCN	SGA-SIM-I	CIRWN	IHS-SCN
AutoMPG8	Training set	RMSE	0.0671	0.0638	0.0639	0.0630	0.0634	**0.0621**
		MAE	0.0486	0.0462	0.0462	0.0454	0.0462	**0.0452**
		R^2	0.8820	0.8945	0.8935	0.8974	0.8959	**0.9005**
	Testing set	RMSE	0.0797	0.0783	0.0771	0.0768	0.0760	**0.0749**
		MAE	0.0582	0.0573	0.0564	0.0562	0.0560	**0.0552**
		R^2	0.8441	0.8571	0.8581	0.8625	0.8675	**0.8688**
Stock	Training set	RMSE	0.0370	0.0318	0.0294	0.0311	0.0316	**0.0275**
		MAE	0.0295	0.0252	0.0232	0.0246	0.0250	**0.0217**
		R^2	0.9744	0.9812	0.9840	0.9820	0.9814	**0.9860**
	Testing set	RMSE	0.0394	0.0350	0.0312	0.0346	0.0347	**0.0297**
		MAE	0.0305	0.0273	0.0241	0.0267	0.0266	**0.0231**
		R^2	0.9702	0.9765	0.9813	0.9770	0.9769	**0.9832**
Concrete	Training set	RMSE	0.0888	0.0782	**0.0769**	0.0783	0.0794	0.0777
		MAE	0.0683	0.0600	**0.0595**	0.0606	0.0612	0.0598
		R^2	0.7592	0.8229	**0.8279**	0.8222	0.8164	0.8254
	Testing set	RMSE	0.1062	0.1040	**0.1010**	0.1038	0.1039	0.1012
		MAE	0.0825	0.0782	0.0774	0.0787	0.0785	**0.0765**
		R^2	0.7170	0.7509	**0.7610**	0.7517	0.7437	0.7595
ANACALT	Training set	RMSE	0.1092	0.0850	0.0795	0.0776	0.0963	**0.0587**
		MAE	0.0797	0.0608	0.0531	0.0479	0.0689	**0.0321**
		R^2	0.7360	0.8551	0.8735	0.8813	0.8054	**0.9356**
	Testing set	RMSE	0.1141	0.0922	0.0864	0.0883	0.1054	**0.0691**
		MAE	0.0814	0.0642	0.0559	0.0506	0.0708	**0.0344**
		R^2	0.7101	0.8301	0.8483	0.8441	0.7709	**0.9080**

Compared with those for the other five models, the evaluation metrics for the IHS-SCN are optimal for the AutoMPG8, Stock, and ANANCALT datasets. For example, for the Stock dataset, the RMSE of the IHS-SCN for the training and testing sets decreased by 13.5% and 15.1%, respectively, compared with those of the SCN. In addition, for the ANANCALT dataset, the RMSE of the IHS-SCN for the training and testing sets decreased by 30.9% and 25.0%, respectively, indicating that using directional derivatives to improve the hidden layer output matrix and the supervisory mechanism through adaptive angle changes can effectively improve the generalization performance and accuracy of the model.

As shown in Table 3, the proposed IHS-SCN performs only slightly worse than the PSCN on the Concrete dataset. The reason behind this may be that the PSCN incorporates an L_1 regularization term in its model construction objective function, which makes

its output weights relatively smaller and sparser. This improves the prediction accuracy of the PSCN model on the Concrete dataset. However, it is worth noting that on the Concrete dataset, the IHS-SCN has better performance than the RVFL, SCN, and CIRWN models in different evaluation metrics. For example, it reduces the RMSE metric by 4.71%, 2.70%, and 2.60% compared with the RVFL, SCN, and CIRWN, respectively. This demonstrates the effectiveness of the improvement strategies proposed in this paper.

To intuitively demonstrate the effectiveness of the IHS-SCN, the training error curves of each model are presented for the four standard datasets in Fig. 3. The figure shows that compared with those of the other five models, the training error curve of the IHS-SCN model is lower during the training process, indicating the effectiveness of improving the hidden layer output matrix and supervisory mechanism.



(a) AutoMPG8 training error curve

(b) Stock training error curve

(c) Concrete training error curve

(d) ANACALT training error curve

Fig.3 Error curve of each model during training.

4.4 Comparison Experiment for Flue Gas Oxygen Content Prediction

In the process of municipal solid waste incineration, the flue gas oxygen content is a key indicator of the incineration efficiency. If the flue gas oxygen content is too high, it indicates that the air volume ratio may be unreasonable, leading to increased heat loss within the incinerator. Conversely, if the flue gas oxygen content is too low, it can result in incomplete combustion within the incinerator, leading to an increase in secondary polluting substances produced by the incineration process[22]. Therefore, establishing an accurate prediction model for the oxygen content in flue gas is of great significance for workers to monitor the incineration status in a timely manner and achieve smooth incinerator operation.

The experimental data consist of 1000 historical records from a solid waste incineration plant in Beijing from December 2021. The parameter settings for each model are shown in Table 2. The experimental results obtained by running each model independently 30 times are shown in Table 4 and Fig. 4.

Table 4 shows that compared with the other five models, the method proposed in this paper yields better

Table 4 Comparison of the experimental results

Dataset	Metric	RVFL	SCN	PSCN	SGA-SIM-I	CIRWN	IHS-SCN
Training set	RMSE	0.0332	0.0261	0.0276	0.0263	0.0254	**0.0235**
	MAE	0.0250	0.0195	0.0204	0.0199	0.0192	**0.0178**
	R^2	0.9730	0.9835	0.9813	0.9833	0.9845	**0.9868**
Testing set	RMSE	0.0368	0.0315	0.0304	0.0313	0.0310	**0.0300**
	MAE	0.0270	0.0223	0.0223	0.0226	0.0225	**0.0211**
	R^2	0.9670	0.9757	0.9774	0.9759	0.9764	**0.9781**

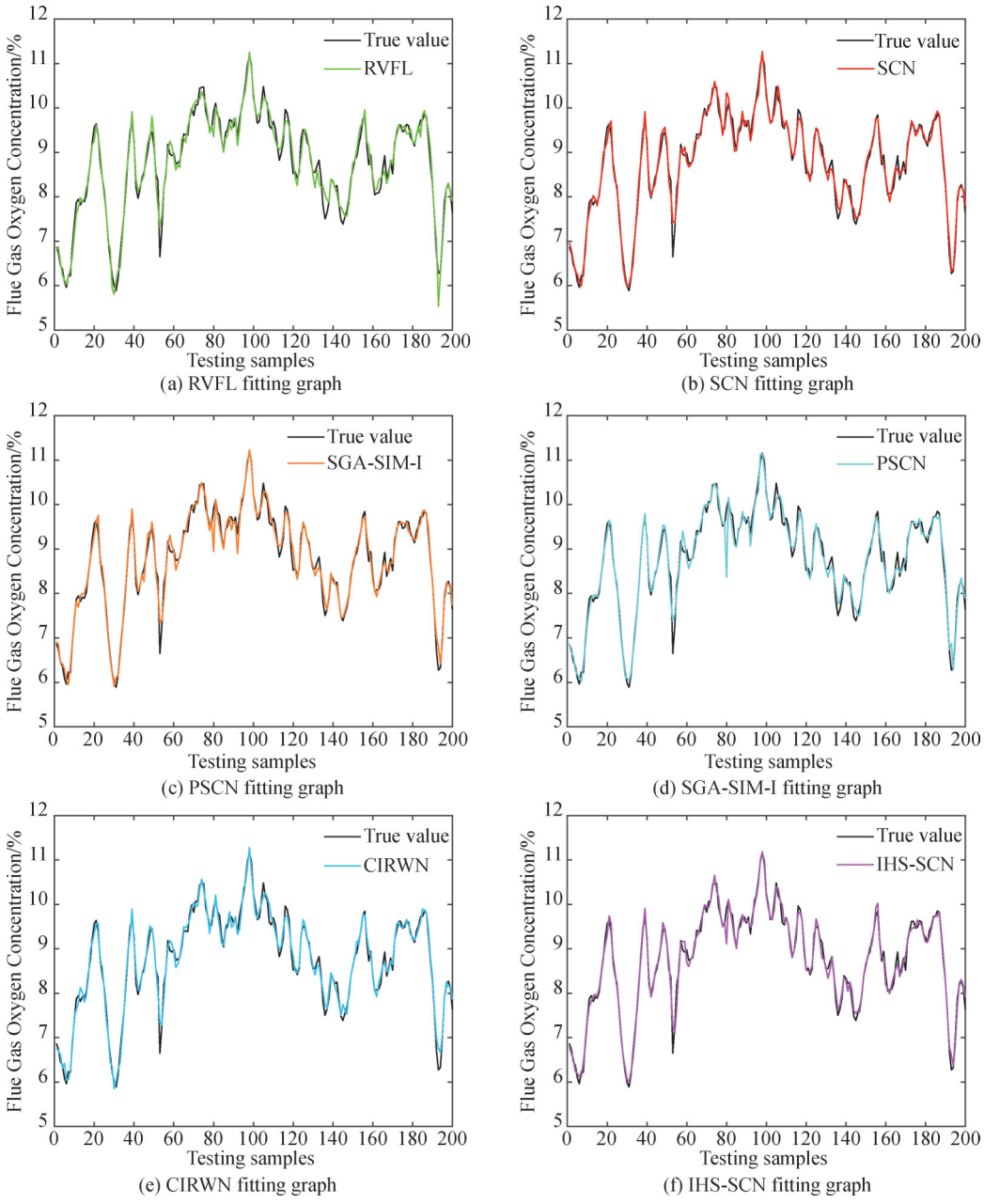


Fig.4 Comparison of the fitting graph for each model

predictive ability. This finding indicates that the new hidden layer output matrix constructed with directional derivatives can effectively improve the generalization ability of the model and that the supervisory mechanism improved via spatial angle adaptation can effectively enhance the prediction accuracy of the model.

Fig. 4 shows the flue gas oxygen content prediction fitting curves for the above models. Notably, compared with that of the RVFL, SCN, PSCN, SGA-SIM-I, and CIRWN, the fitting effect of the method proposed in this

paper is the best, verifying the effectiveness of the proposed improvement method.

5 Conclusion

The modeling method based on an improved hidden layer output matrix and supervisory mechanism (IHS-SCN) is developed to improve the prediction accuracy and generalization performance of the SCN model, and the effectiveness of the IHS-SCN method is verified through experiments. The main contributions of this study are as follows.

First, a new hidden layer output matrix is constructed. By calculating the first and second directional derivatives of the original hidden layer output matrix and linearly adding the key factors affecting the generalization performance of the model to the original hidden layer output matrix, a new hidden layer output matrix is obtained; this approach enhances the mapping ability and generalization performance of the model in different modeling tasks.

Second, a new supervisory mechanism inequality is proposed. The new supervisory mechanism adaptively changes the constraint degree considering the spatial angle variation between the residuals and activation function, thus improving the quality of the hidden layer parameters, avoiding the problem of low quality for newly added nodes in the model due to the limits of constraints, and improving the modeling accuracy.

The experimental results indicate that the proposed IHS-SCN model yields high accuracy and generalization ability. Industrial data experiments show that the flue gas oxygen content prediction model based on the IHS-SCN can accurately predict flue gas oxygen content changes and has certain application advantages over other models. Owing to the need for pseudoinverse operations involving the output matrix during modeling and the high dimensionality of the output matrix in the method proposed in this paper, the level of resource consumption is high when the pseudoinverse is calculated. Therefore, simplifying the pseudoinverse operation of the output matrix will be the focus of subsequent research. Additionally, the prediction accuracy of the improved SCN model is not ideal for some datasets, and it is necessary to explore in detail the characteristics of various datasets to further enhance model performance.

Author Contribution:

Aijun Yan: Conceptualization; Data curation; Funding acquisition; Supervision; Writing-review & editing. Jing Wang: Investigation; Methodology; Software; Validation; Visualization; Writing-original draft.

Foundation Information:

This work is supported by the National Natural Science Foundation of China (62373017).

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Conflicts of Interest:

The authors declare no competing interests.

Dates:

Received 17 July 2025; Accepted 17 December 2025; Published online 31 December 2025

References

[1] Yin S, Ding S X, Xie X, et al., "A review on basic data-driven approaches for industrial process monitoring, " *IEEE Transactions on Industrial Electronics* , vol. 61, no. 11, pp. 6418-6428, Jan. **2014**.

[2] Lopez-Garcia T B, Coronado-Mendoza A, Domínguez-Navarro J A, "Artificial neural networks in microgrids: A review," *Engineering Applications of Artificial Intelligence* , vol. 95, no. 8, pp. 103894, Aug. **2020**.

[3] Jia R, Zhang S, You F, "Nonlinear soft sensor development for industrial thickeners using domain transfer functional-link neural network," *Control Engineering Practice* , vol. 113, no. 8, pp. 104853, Aug. May **2021**.

[4] Scardapane S, Wang D H, "Randomness in neural networks: an overview," Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery* , vol. 7, no. 2, pp. e1200, Apr. **2017**.

[5] Wang D H, "Randomized algorithms for training neural networks," *Information Sciences* , vol. 100, no. 364-365, pp. 126-128, **2016**.

[6] Pao Y H, Park G H, Sobajic D J, "Learning and generalization characteristics of the random vector functional-link net, " *Neurocomputing* , vol. 6, no. 2, pp. 163-180, Dec. **1994**.

[7] Ren Y, Suganthan P N, Srikanth N, et al., "Random vector functional link network for short-term electricity load demand forecasting, " *Information Sciences* , vol. 367, no. 11, pp. 1078-1093, Nov. **2016**.

[8] Yu L, Wu Y, Tang L, Yin H, Lai K K, "Investigation of diversity strategies in RVFL network ensemble learning for crude oil price forecasting," *Soft Computing* , vol. 25, no. 10, pp. 3609-3622, Oct. **2021**.

[9] Li M, Wang D H, "Insights into randomized algorithms for neural networks: Practical issues and common pitfalls, " *Information Sciences* , vol. 382, no. 3, pp. 170-178, Dec. **2017**.

[10] Wang D H, Li M, "Stochastic configuration networks: fundamentals and algorithms, " *IEEE Transactions on Cybernetics* , vol. 47, no. 10, pp. 3466-3479, Sep. **2017**.

[11] Wang Q J, Dai W, Ma X P, et al., "Driving amount based stochastic configuration network for industrial process modeling," *Neurocomputing* , vol. 394, no. 2, pp. 61-69, Feb. **2020**.

[12] Dai W, Li D, Chen Q, et al., "Data driven particle size estimation of hematite grinding process using stochastic configuration network with robust technique, " *Journal of Central South University* , vol. 26, no. 1, pp. 43-62, Apr. **2019**.

[13] Yan A J, Guo J C, Wang D H, "Heterogeneous feature ensemble modeling with stochastic configuration networks for predicting furnace temperature of a municipal solid waste incineration process, " *Neural Computing and Applications* , vol. 34, no. 18, pp. 15807-15819, Apr. **2022**.

[14] Wang Q J, Dai W, Lu Q, et al., "A sparse learning method for SCN soft measurement model, " *Control and Decision* , vol. 37, no. 12, pp. 3171-3182, Aug. **2021**.

[15] Wang Q J, Yang C Y, Ma X P, et al., "Underground airflow quantity modeling based on SCN," *Acta Automatica Sinica* , vol. 47, no. 8, pp. 1963-1975, Nov. **2021**.

[16] Zhu X, Feng X, Wang W, et al., "A further study on the inequality constraints in stochastic configuration networks, " *Information Sciences* , vol. 487, no. 3, pp. 77-83, Mar. **2019**.

[17] Hu K C, Yan A J, Tang J, "Multi-target robust prediction model for furnace temperature and flue gas oxygen content in municipal solid waste incineration process," *Acta Automatica Sinica* , vol. 50, no. 5, pp. 1001−1014, Jan. **2024**.

[18] Nan J, Dai W, Yuan G, et al., "A stochastic incremental learning model with maximizing spatial geometry angle and its application, " *Acta Automatica Sinica* , vol. 49, no. 6, pp. 1283-1294, Sep. **2023**.

[19] Huang C Q, Li M, Wang D H, "Stochastic configuration network ensembles with selective base models, " *Neural Networks* , vol. 137, no. 1, pp. 106-118, Jan. **2021**.

[20] Dattorro J, "Appendix D Matrix Calculus, " *in Convex optimization and Euclidean distance geometry* , California, USA: Meboo, 2010, pp: 621-650.

[21] Wang Q J, Dai W, Lin P, et al., "Compact incremental random weight network for estimating the underground airflow quantity," *IEEE Transactions on Industrial Informatics* , vol. 18, no. 1, pp. 426-436, Mar. **2021**.

[22] Vilardi G, Verdone N, "*Exergy analysis of municipal solid waste incineration processes* : The use of O2-enriched air and the oxy-combustion process, " Energy, vol. 239, no. 9, pp. 122147, Sep. **2021**.