Current Science
CS 5 (2), 2063-2082 (2025)
**Received 15 Feb** 2025 | **Revised 19 March** 2025 | **Accepted 12 April**. 2025 | **Online Available 26 April** 2025
https://doi.org/ 10.52845/CS/2025-5-2-63

Current Science

Current Science

OPEN ACCESS

ISSN(O)2795-8639

**Research Article**

# Short-Term Wind Speed Prediction Model Based on Secondary Decomposition and SE-SSA-TCN

**Guo Qi[1], Wang Wenbo[1]**

**[1]College of science, Wuhan University of Science and Technology, Wuhan 430000, China**

**[*]Corresponding Author: Guo Qi**

**Abstract:**

**Background:** Aiming to enhance the accuracy of wind speed prediction, a hybrid framework (secondary decomposition-sample entropy-sparrow search algorithm-time convolutional network) was proposed.

**Methods:** This framework integrates SD and SE-SSA to optimize the TCN model. Additionally, a complete ensemble empirical mode decomposition with adaptive noise-variational mode decomposition mechanism was developed to address signal decomposition challenges. A sample entropy strategy was employed to reduce model complexity while maintaining predictive performance. The Sparrow Search Algorithm was utilized to overcome the limitations of traditional optimization methods.

**Results And Conclusions:** Experimental results based on US power station data demonstrated that the proposed model significantly reduced RMSE, MAE, and MAPE compared to single TCN and SD-SE-TCN models, thereby validating its high accuracy and excellent generalization capability.

**Keywords: Short-term wind speed forecasting; secondary decomposition; time convolutional network; sparrow search algorithm; sample entropy-based recombination**

## Introduction

With the acceleration of the global energy structure adjustment process, wind energy has become an important alternative energy to cope with the environmental crisis because of its cleanliness and renewability [1].However, under the influence of complex meteorological factors such as atmospheric turbulence and temperature gradient, wind speed series presents significant non - stationarity and multi-scale characteristics [2], resulting in a large deviation in wind power prediction. According to the statistics of the US Department of Energy, every 1% reduction in the prediction error can reduce the system standby capacity by 10%-15%[3], which makes high-precision prediction technology the key to ensure the stable operation of the power grid.

The existing forecasting methods mainly use the "decomposition and reconstruction" framework to alleviate the non - stationarity problem. Typical decomposition algorithms include empirical mode decomposition (EMD)[4], wavelet transform (WT)[5] and variable mode decomposition (variable mode decomposition). VMD)[6] and so on. Literature [7] adopts Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose wind speed series. However, mode aliasing still exists in the residual modes; Literature [8] improved the prediction performance by integrating ensemble empirical mode decomposition (EEMD) - Convolutional neural network (CNN) - XGBoost combined model, but ignored the secondary noise interference of high-frequency components. The results show that some signal subsequences still have significant noise and high frequency after a single decomposition, which indicates that it is difficult to fully extract hidden features from a single decomposition. In view of this, the CEEMDAN - VMD collaborative secondary decomposition strategy for specific sub-signals may achieve deep stripping of residual noise and significantly improve the accuracy of short-term wind speed prediction [9].

In the field of deep learning, convolution neural networks (CNN)[10] and long short term memory (LSTM)[11] have made outstanding contributions. Time Convolutional Network (TCN) [12], developed on the basis of CNN, overcomes the limitations of CNN, such as difficulty in using time series sequence information and lack of long-distance dependent modeling ability, by means of the characteristics of causal convolution, extended convolution and multi-layer stacking. In literature [13], (Bidirectional LSTM) BiLSTM was used to enhance bidirectional dependency modeling ability, but it failed to effectively identify key feature weights. Literature [14] points out that if the secondary decomposition adopts the same algorithm, mode aliasing will be introduced, resulting in modal redundancy accumulation, and the hierarchical convolution of TCN can adaptively filter noise. Although the LSTM - TCN hybrid model constructed in literature [15] can capture the temporal features, it has the problem of gradient disappearance.

In the field of deep learning, the actual performance of a model is often directly related to the value of hyperparameters [16].In literature [17], grey wolf optimizer (GWO) was used to optimize the number of hidden layer units and learning rate of LSTM prediction model to further reduce the prediction errors. However, the traditional optimization algorithm is easy to fall into local optimal. In recent years, Sparrow Search Algorithm (SSA) proposed by Xue et al. [18] can carry out global search and local development by simulating the interaction between finder and entrant in sparrow foraging process and sparrow's response mechanism in the face of danger. The dynamic balance between global exploration and local development is realized [19], and the global optimal solution of the problem can be searched more effectively.

To solve the above problems, a hybrid architecture of secondary decomposition (CEEMDAN - VMD) - Sample Entropy (SE) - Sparrow Search Algorithm (SSA) - Time Convolutional Network (TCN) timing prediction is constructed in this study. Firstly, a secondary decomposition mechanism is designed. CEEMDAN is used to achieve coarse signal decomposition, and VMD is used to reconstruct the noise-containing component ($IMF_1$).Secondly, sample values are obtained by using sample entropy, and then subsequences with similar complexity are combined to form new sequences to reduce the number of prediction models [20].Then, SSA was used to optimize TCN, efficiently adjust parameters, avoid local optimizations, accelerate convergence, deeply explore the nonlinear and time series characteristics of wind speed data, improve model prediction accuracy and generalization ability, and provide more reliable support for short-term wind speed prediction. Finally, in order to verify the validity of the above method, the measured data of a power station in the United States are collected and analyzed.

## 1 Secondary Decomposition

Complex meteorological conditions and other factors have a great impact on wind speed series, which makes it show significant volatility and randomness, and the accuracy is difficult to guarantee when directly forecasting it. In order to improve the accuracy and stability of short-term wind speed prediction, this paper adopts the secondary decomposition method to process the original wind speed data. Firstly, the original wind speed signal is decomposed into several eigenmode function (IMF) subsequences with different characteristic scales using CEEMDAN algorithm. Secondly, VMD is used to construct and solve the variational model, and $IMF_1$ is further decomposed into multiple modal components with narrower frequency bands, effectively reducing the non - stationarity and nonlinear degree of the high-frequency component.

### 1.1 Ceemdan

In view of the mode aliasing problem existing in EMD, Torres et al. [20] proposed CEEMDAN. This method inherits the advantages of ensemble empirical Mode decomposition (EEMD)[21] and overcomes the problems of difficult noise elimination and low decomposition efficiency through the synergistic effect of adaptive white noise injection mechanism and ensemble averaging strategy. Let the KTH modal component after decomposition of EMD algorithm be $E_K(i)$, and the KTH component generated by CEEMDAN algorithm be $IMF_K$. The detailed steps of the CEEMDAN algorithm are as follows.

(1) K sets of independent white noise sequences

(mean 0, variance 1) $\omega_i(t)$ are generated, plus and minus symmetric white noise is superimposed on the one-dimensional signal $x(i)$ waiting for decomposition, and two sets of disturbance signals are constructed to cancel the residual noise.

$$x_i^+(t) = x(t) + \varepsilon\omega_i(t) \quad (1)$$

$$x_i^-(t) = x(t) - \varepsilon\omega_i(t) \quad (2)$$

Where, ε is the noise amplitude coefficient; The standard deviation of $\omega_i(t)$ is proportional to the signal standard deviation; $i = 1, 2, \cdots, K$ .

(2) Each group $x_i^+(t)$ and $x_i^-(t)$ were decomposed by EMD, and the first order modal component $IMF_1$ was obtained by averaging the K first components of the decomposition, and the first-order IMF extraction was realized, and the first-order residual component $x_i^-(t)$ was obtained.

$$IMF_1(t) = \frac{1}{2K}\sum_{i=1}^{K}(EMD_1(x_i^+(t) + EMD_1(x_i^-(t))) \quad (3)$$

$$r_1(t) = x(t) - IMF_1(t) \quad (4)$$

(3) For the J-order IMF ( $j = 2, 3, \cdots, K$ ), adaptive noise is superimposed to the residual signal of the J-stage obtained after decomposition to generate a new disturbance signal, and then EMD algorithm is used for decomposition to extract and average the first-order IMF. Update the residual at the same time.

$$IMF_j(t) = \frac{1}{K}\sum_{i=1}^{K}E_1[r_{j-1}(t) + \varepsilon_{j-1}E_{j-1}(\omega_i(t))] \quad (5)$$

$$r_j(t) = r_{j-1}(t) - IMF_j(t) \quad (6)$$

(4) Repeat step (3) to terminate the decomposition when the residual signal becomes a monotone function or the number of extremum points does not exceed 2, and finally obtain the CEEMDAN decomposition sequence:

$$x(t) = \sum_{n=1}^{K}IMF_n + R(t) \quad (7)$$

The original wind speed time series was adaptively decomposed into nine intrinsic mode functions (IMFs) through CEEMDAN algorithm. As visually presented in Figure 1, the decomposition yields distinct wind power subsequences with differentiated temporal characteristics. Critical examination of the decomposed components reveals that CEEMDAN effectively mitigates the inherent nonlinearity and stochasticity embedded in the primitive wind speed signal, as quantitatively verified by subsequent entropy metrics and stationarity tests.

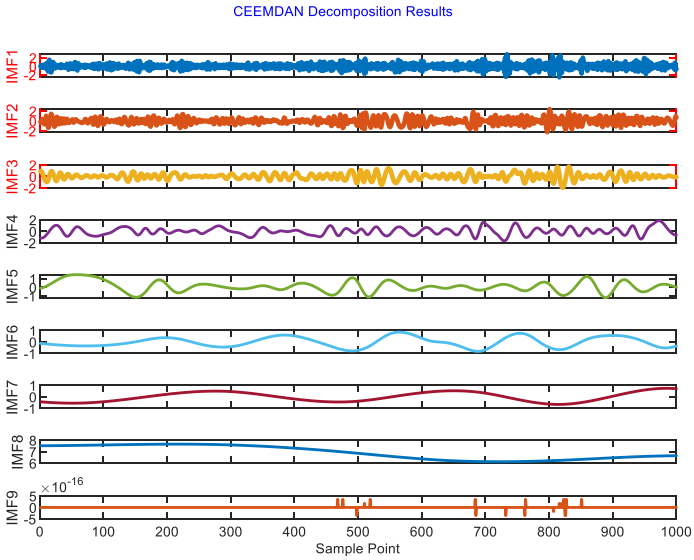


**Figure 1 Wind speed sequence after CEEMDAN decomposition**

### 1.2 VMD

In this paper, variational mode decomposition (VMD) was used to decoupage the secondary signals of $IMF_1$ from the decomposed CEEMDAN.

Based on the principle of adaptive band segmentation, VMD achieves accurate division of frequency domain by constructing constrained variational variants, which can better suppress band overlap and has convergence provability. In other words, the variational model is solved by alternating direction multiplier method (ADMM) to ensure that the decomposition process converges to the global optimal solution [22].

The VMD algorithm is:

$$\min_{\{v_k\},\{w_k\}}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-i\omega_k t}\right\|_2^2\right\} \quad (8.)$$
$$s.t.\sum_k u_k = IMF_1$$

Where, $u=\{u_1,u_2,\cdots u_K\}$ represents K eigenmode functions (BLIMF) after IMF1 is decomposed, $\{\omega\}=\{\omega_1,\omega_2,\cdots,\omega_K\}$ is the frequency center of each component, and $\partial_t$ represents the time derivative operator. By introducing quadratic penalty factor α and Lagrange multiplier λ(t), the above problem can be transformed into an unconstrained optimization form for iterative solution [23]. The extended Lagrange expression is as follows:

$$L(\{u_k\},\{\omega_k\},\lambda)=\alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)u_k(t)\right]e^{-i\omega_k t}\right\|_2^2+$$
$$\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle \quad (9)$$

As shown in Figure 2, after the secondary decomposition of the VMD, the original IMF supply forms were decoupled into five narrow strip sequences (BLIMFs) of clear physical significance. According to the calculation of Figure 3, the average sample entropy of the decomposed sub-sequence is 0.24, and that of the original IMF1 sequence is 1.65. The average sample entropy of the decomposed sub-sequence is reduced by 85.03%, which significantly improves the input quality of the subsequent prediction model.

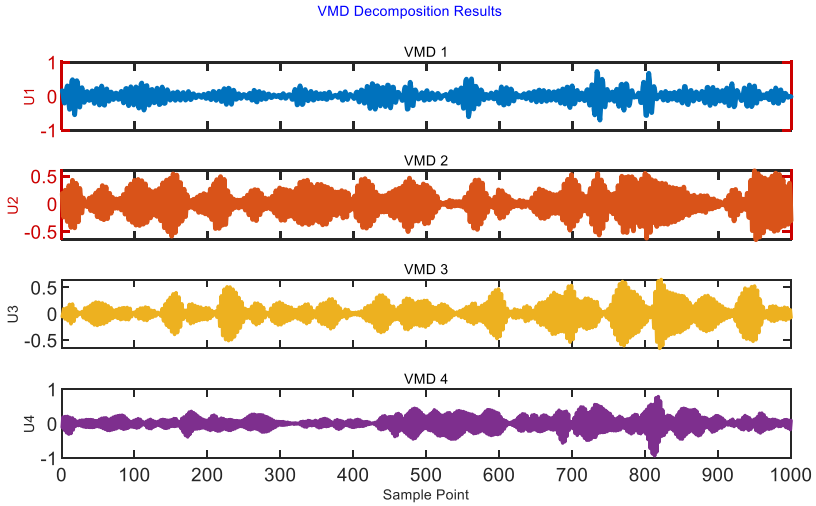



**Figure 2 $IMF_1$ sequence after VMD decomposition**

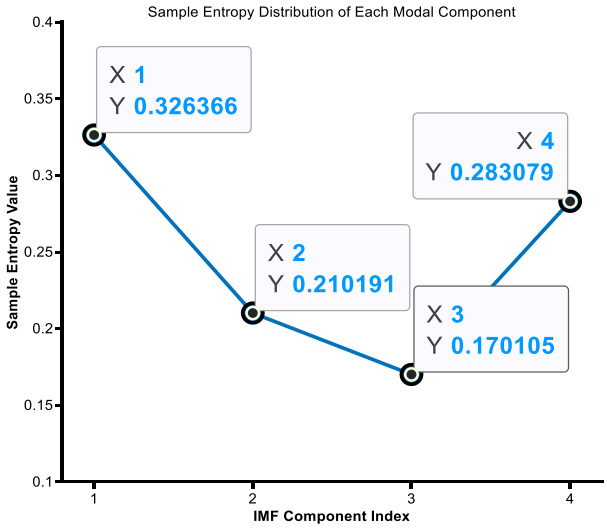



**Figure 3 Entropy of VMD decomposition subsequence**

## 2 Establishment of Wind Speed Prediction Model

### 2.1 Modal Recombination Optimization Based on Sample Entropy

As a quantitative index of time series complexity, sample entropy (SE) has the core advantage of measuring the robustness of pattern repetition probability [24]. Compared to approximate entropy (ApEn)[25], SE significantly improves stability by eliminating self-matching counts (and maintaining parameter consistency). Given the submode set after CEEMDAN-VMD secondary decomposition, the SE calculation process is defined as follows:

(1) SE was defined, submode sequence $x=\{x(1),x(2),\cdots,x(n)\}$ was set, dimension m was defined, and similarity tolerance r was defined. m=2 was adopted to balance computational efficiency and feature sensitivity. Standard deviation ratio method was adopted, and r=0.2σ (σ is sequence standard deviation) was set [26].

(2) Phase space reconstruction:

$$X_m(i)=[x(i),x(i+1),\cdots,x(i+m-1)] \quad (10.)$$

In the formula, $i=1,2,\cdots,N-m+1$.

(3) Calculate the pattern matching probability:

$$B^m(r)=\frac{1}{N-m}\sum_{i=1}^{N-m}\frac{\sum_{j=1,j\neq i}^{N-m}\Theta(r-d[X_m(i),X_m(j)])}{N-m-1} \quad (11.)$$

(4) Where $\Theta(\cdot)$ is the Heaviside step function and $d[\cdot]$ is the Chebyshev distance.

In practical application, the sequence length N is a finite value, then the estimated value of SE is:

$$SampEn(m,r,N)=-\ln\left[\frac{B^{m+1}(r)}{B^m(r)}\right] \quad (12)$$

The smaller the value of the sample entropy corresponding to the time series, the more regular and predictable it is. On the contrary, the larger the value of sample entropy, the more complex and random it is [27]. The sample entropy establishes the screening criteria for the submode sequence, which can effectively eliminate the noise modes with high SE value. The calculation results are shown in Figure 3. If the sample entropy of sequence IMF1 is greater than 1.6, it indicates that it is highly complex and difficult to predict, so it needs to be discarded. With 0.2 as the demarcation standard, the SE values of $IMF_3$ and $IMF_4$ were similar. SE values of $IMF_6$, $IMF_7$, $IMF_8$ and $IMF_9$ are between 0.2 and 0. The SE values of sequences $IMF_2$ and $IMF_5$ are far from the same standard as those of other sequences. Sequences with similar SE values were superimposed to form 4 superimposed components, as shown in Figure 4. As the input of

SSA - TCN model, the reconstituted signal complexity is reduced, which is conducive to improving the timing modeling accuracy of TCN network and adapting the prediction model. Figure 5 shows the original sequence superimposed to form the new component, which will become the input sequence.

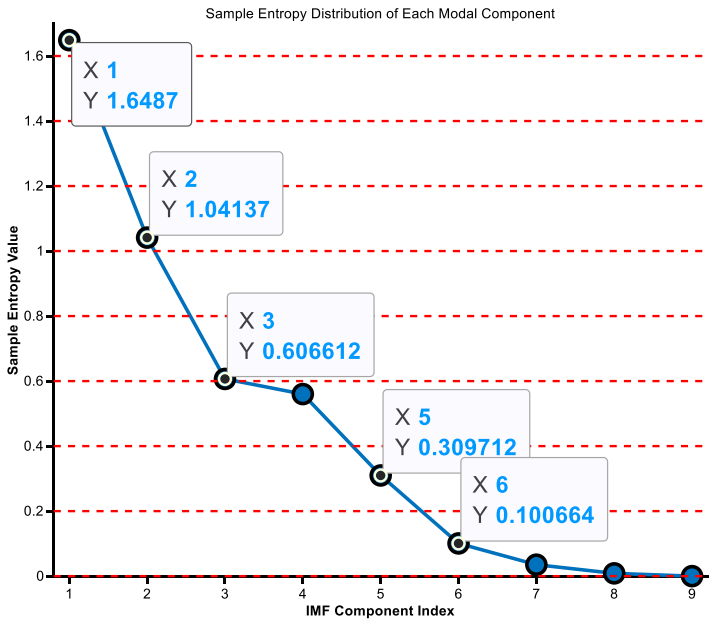



**Figure 4 Sample entropy values of each sequence after CEEMDAN decomposition**

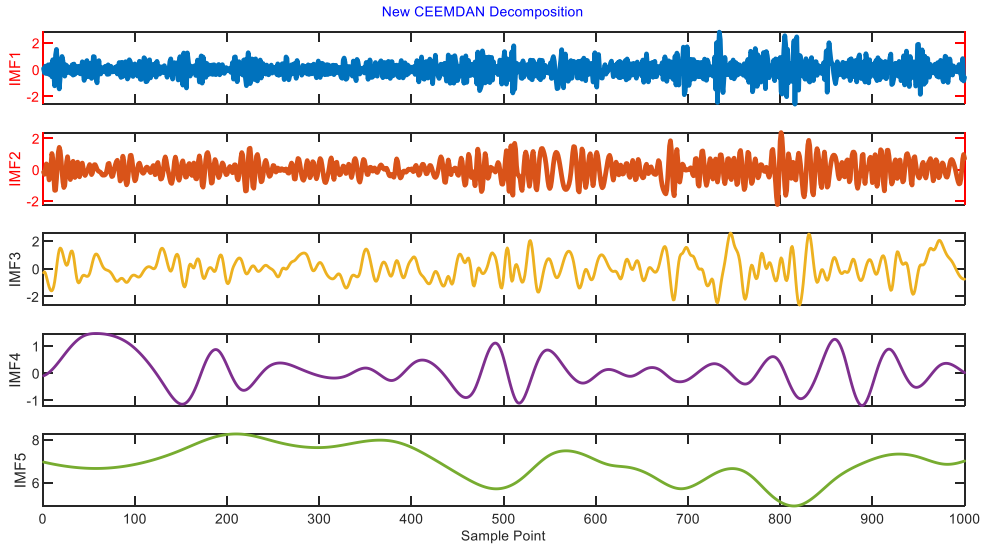



**Figure 5 New subsequences after stacking according to SE values**

## 2.2 TCN Model

Time convolutional network (TCN), with its unique dilatative causal convolution and residual learning mechanism, shows significant advantages in complex time series prediction tasks [28]. In this study, TCN is used as the core prediction module to process the wind speed submode signal reconstructed by CEEMDAN-VMD secondary decomposition and sample entropy (SE) screening. Compared with traditional recurrent neural networks such as LSTM, TCN has better performance in long-term dependence modeling, training efficiency and noise robustness. The structure of the TCN prediction model constructed in this research is shown in Figure 6, and its core components and mathematical characterization are as follows:

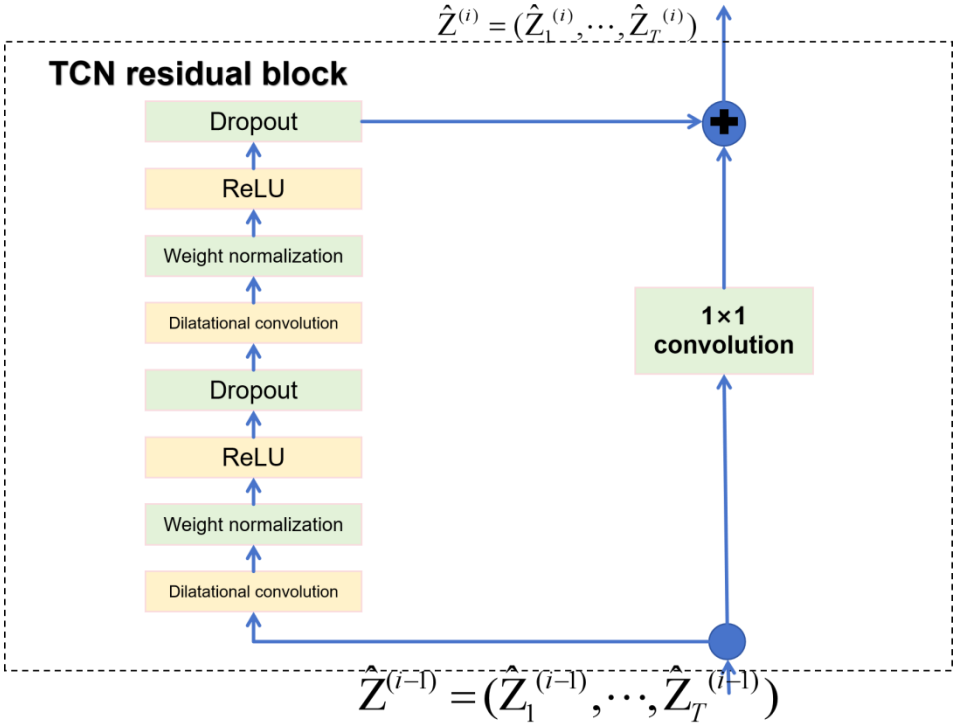



**Figure 6 TCN residual block**

(1) In the dilative causal convolution layer, the sequence $X \in \mathbb{R}^{L \times d}$ after modal recombination (L is the sequence length, d is the feature dimension) and the dilative convolution operation are defined as:

$$Z_t^{(l)} = \sum_{k=0}^{K-1} W_k^{(l)} \cdot X_{t-k \cdot D^{(l)}} \qquad (13)$$

expansion coefficient of layer l, which grows exponentially to cover long range historical information. K is the convolution layer size, $W_k^{(l)}$ is the learnable weight matrix, and weight normalization is used to accelerate convergence. The sub-modes decomposed by CEEMDAN - VMD contain wave characteristics of different time scales, and the multi-scale sensitivity field of expansion convolution can explicitly model the wind speed change mode from minute level to day level [29]. And through causal constraint $(t - k \cdot D^{(l)} \le t)$, TCN ensures that the prediction relies only on historical data and meets the temporal causality requirements of wind speed prediction.

(1) Residual joins are introduced after each convolution layer to mitigate the problem of disappearing gradients:

$$O^{(l)} = ReLU(Z^{(l)} + F(X^{(l)}) \qquad (14)$$

$F(\cdot)$ is a 1×1 convolution used to adjust the channel dimension to ensure that the residual path matches the main path dimension. The submodes after sample entropy screening may still retain low frequency noise, and the hierarchical convolution structure of TCN has the capability of progressive feature noise reduction: shallow convolution captures high frequency details, and deep convolution extracts low frequency trends.

(2) TCN combines abstract features of different levels by jumping connections to achieve multi-scale feature fusion.

$$H_{final} = \sum_{l=1}^{L} \alpha_l \cdot Z^{(l)} \qquad (15)$$

In the formula, $\alpha_l$ is the learnable attention weight, which can be optimized and initialized by the optimization algorithm to dynamically adjust the contribution degree of features at all levels.

### 2.3 SSA

Sparrow search algorithm (SSA), proposed by scholars Xue et al. [30], builds an efficient swarm intelligence optimization framework by simulating the hierarchical foraging strategy and

dynamic hedging mechanism of sparrow population. The algorithm divides population individuals into three types of roles. The first is the discoverer mechanism, which accounts for 20%-30% of the population size and is responsible for exploring high-yield areas with globally optimal fitness values. Its energy reserve level is positively correlated with fitness value:

$$E_i = \frac{f_{\max} - f_i}{\sum_{j=1}^{N}(f_{\max} - f_j)} \tag{16}$$

When the warning coefficient $R_2 \geq ST$ (safety valve value ST∈[0.5,1]), the discoverer triggers the hedging behavior and guides the population to migrate to the safe area. The secondary operational phase employs a follower-dominant paradigm, where the follower subpopulation comprises 60-70% of the swarm population and engages in intensified local exploitation. The position updating mechanism is governed by dual evolutionary pressures: attraction forces emanating from dominant discoverers and intra-population competitive dynamics. Notably, stochastic competition is mathematically formalized through the strategic integration of Cauchy mutation operators C(0,1), which probabilistically simulate resource competition mechanisms by inducing controlled perturbations in the solution space, thereby enhancing swarm diversity while maintaining convergence properties. The third is the alert behavior, which accounts for 10%-20% and is distributed at the edge of the population. Levy flight strategy is used to explore the heavy-tail distribution of escape paths:

$$X_s^{t+1} = X_{best}^t + \gamma \cdot Levy(\beta) \otimes (X_s^t - X_{best}^t) \tag{17}$$

Where γ is the step scaling factor and Levy(β) is a random vector of Levy distribution obeying the parameter β. At the same time, SSA allows individuals to transfer roles according to fitness changes. When the follower's continuous k generation fitness improvement rate is below the threshold δ, it converts to an alert. If the alert finds a new optimal solution, it is promoted to discoverer. This mechanism maintains population diversity through elite retention strategy and inferior elimination mechanism. The construction of SSA model consists of the following steps:

Firstly, the population is initialized, N sparrow individuals are generated in the solution space, and the initial fitness $f_i$ is calculated. Then the roles are assigned, sorted according to fitness, the first $N_p$ is the discoverer, the last $N_s$ is the alert, and the rest are the followers. Then the finder performs wide-area exploration according to formula (17), the follower performs local development according to formula (18), and the alert performs escape search according to formula (19) for iterative optimization.

$$X_p^{t+1} = \begin{cases} X_p^t \cdot \exp(-\dfrac{t}{\eta T_{\max}}), R_2 < ST \\ X_p^t + \sigma \cdot L \cdot \aleph(0,1), R_2 \geq ST \end{cases} \tag{18}$$

$\eta \in (0,1)$ is the attenuation coefficient, σ is the intensity of the Gaussian disturbance, and L is the D-dimensional unit direction vector.

$$X_f^{t+1} = \begin{cases} X_p^t + A^+ \cdot (X_f^t - X_p^t), f_f \geq \overline{f} \\ X_{worst}^t + C(0,1) \otimes (X_f^t - X_{worst}^t), f_f < \overline{f} \end{cases} \tag{19}$$

$A^+$ is the Moore-Penrose pseudo-inverse matrix, and $\overline{f}$ is the population average fitness.

$$X_s^{t+1} = X_s^t + \lambda \cdot \mathrm{sgn}(J) \cdot (\frac{f_s - f_{worst}}{|f_{best} - f_{worst}| + \varepsilon}) \tag{20}$$

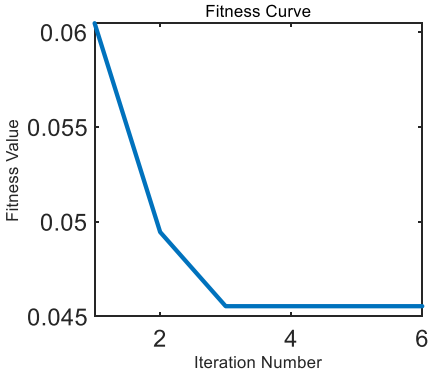
Where λ is an adaptive step, J∈[−1,1] is a random direction factor, and ε is a small constant that prevents division by zero. Terminates when the maximum number of iterations is reached or the optimal solution is stable. The sensitivity analysis of SSA is conducted, and the results are shown in Table 1. It can be seen that the sensitivity of SSA is low, and the alarm mechanism has an optimal triggering frequency.

**Table S1 Sensitivity analysis of SSA parameters**

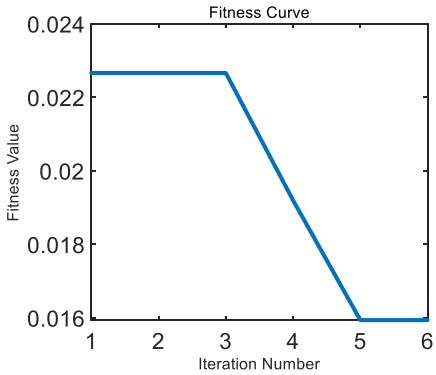
| Argument | Value range | Optimal value | Fitness fluctuation rate |
|---|---|---|---|
| *N* | [5,10] | 8 | ±2.1% |

| | | | |
|---|---|---|---|
| *ST* | [0.5,1.0] | 0.75 | ±1.8% |
| $\eta$ | (0,1] | 0.35 | ±3.2% |

In order to verify that SSA has a strong parameter seeking ability, the fitness curve of SSA is given during the process of TCN's prediction variables [31]. Here, the prediction variables $BLIMF_2$ and $BLIMF_4$ are taken as examples.




(a) $BLIMF_2$ fitness curve




(b) $BLIMF_4$ fitness curve

**Figure 7 SSA fitness curve**

## 2.4 MSSA

In the optimization process of SSA, the convergence speed is slow or the SSA falls into the local optimal problem, which leads to the low efficiency of the experiment process and the optimal parameter configuration of the TCN model. In this paper, adaptive mechanism and hybrid strategy are used to improve its optimization performance [31], and the improved Algorithm is named Modified Sparrow Search Algorithm (MSSA). The Tent chaotic map is first used to generate the initial population to enhance population diversity and ergodicity. Then the ratio of finder and follower is dynamically adjusted according to the iteration progress, and Levy flight is introduced in the finder position update to enhance the global search capability. The specific formula is updated as follows:

(1) Chaotic initialization population:

$$x_{i,j}=\begin{cases}2x_{i,j}, 0\le x_{i,j}\le 0.5\\ 2(1-x_{i,j}), 0.5<x_{i,j}\le 1\end{cases} \quad (21)$$

(2) Adaptive role scaling:

$$N_p(t)=N\cdot(0.3-0.2\cdot\frac{t}{T_{\max}}) \quad (22)$$

(3) Hybrid Levy flight strategy:

$$X_p^{t+1}=X_p^t+\alpha\cdot Levy(\beta)\otimes(X_p^t-X_{best}^t) \quad (23)$$

The adaptive parameters avoid the deviation of manual parameter adjustment, the chaotic initialization increases the population diversity, and the hybrid strategy reduces the probability of premature convergence.

## 2.5 TCN Timing Model Based on SSA Optimization

The SSA - TCN model optimizes the hyperparameters of the time convolutional network by Sparrow search algorithm, and constructs a decomposing, optimization - prediction closed-loop architecture. Its core innovation is to deeply integrate the global optimization capability of SSA with the advantages of multi-scale time series modeling of TCN. The prediction accuracy of TCN network is mainly related to delay step size, learning rate, number of convolution kernel and regularization coefficient [32]. The fitness function is used to evaluate the advantages and disadvantages of each group of chromosomes. In order to determine a set of network hyperparameters that minimize the prediction error of TCN, the fitness function uses the minimum mean squared error between the expected output and the actual output. The model flow of SSA - TCN is shown in Figure 8. The specific steps are as follows:

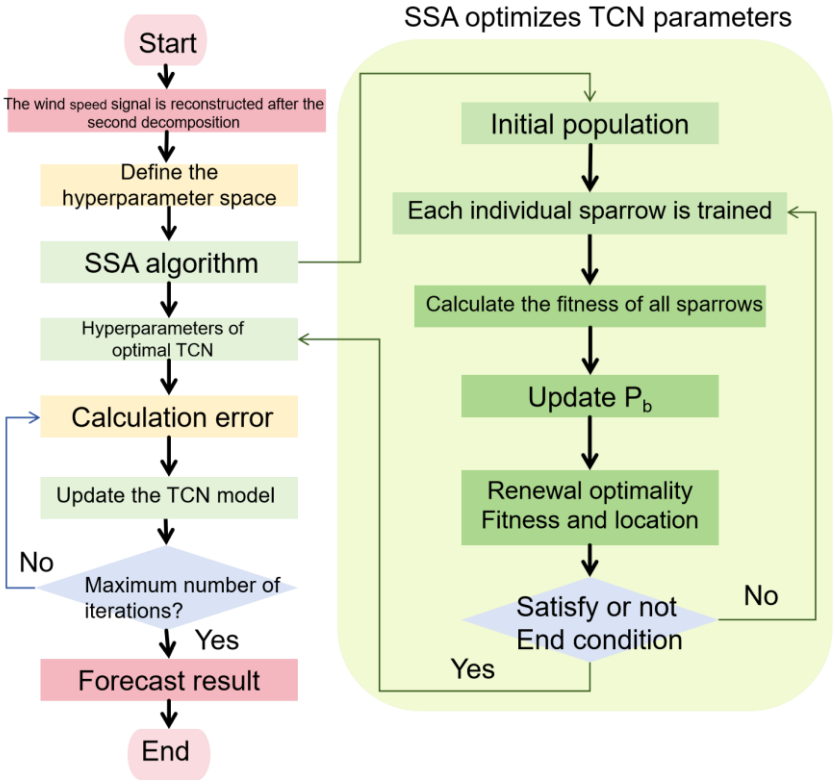



**Figure 8 Flowchart of the SSA-TCN model**

(1) Based on CEEMDAN - VMD quadratic decomposition and sample entropy sub-screening of the reassembled modal signal $\{\mathrm{BLIMF}_k\}_{k=1}^{K}$, a multidimensional input matrix is constructed.

(2) Initialize SSA parameters and TCN network hyperparameters.

(3) The parameters of the current TCN network were used to predict the input data, and the fitness values of all sparrows were calculated.

(4) Update the location of all sparrows according to the rules of the sparrow search algorithm. The current optimal solution and its corresponding fitness value are determined by comparing the fitness values of the updated sparrow individuals.

(5) Repeat the above steps until t satisfies the termination condition. If the termination condition is reached, the global optimal solution is output, and the TCN network parameters corresponding to the global optimal solution are the parameters optimized by SSA. Otherwise, return to Step (4) to continue with the iterative update.

(6) The optimal network hyperparameters are substituted into TCN network for prediction.

2.6 Short-term Wind Speed Prediction Model

Considering the randomness, fluctuation and nonlinear waiting characteristics of wind speed, a wind speed prediction model based on quadratic decomposition, sample entropy and SSA - TCN is proposed in this paper. Figure 9 shows the flow of the wind speed prediction model, with specific steps as follows:

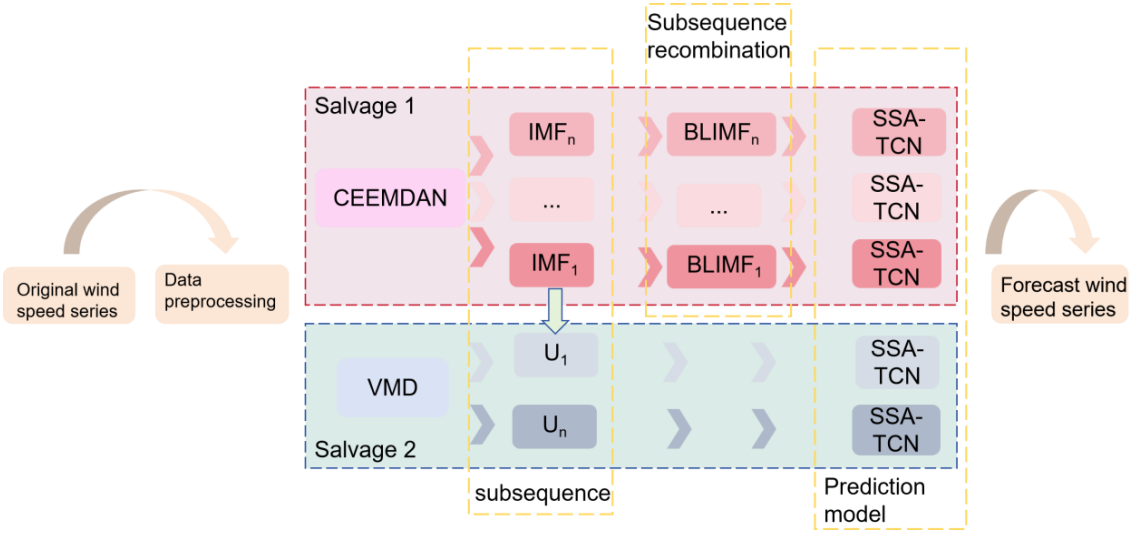



**Figure 9 Flow chart of wind speed prediction model**

(1) Data preprocessing: For the original wind speed data, abnormal data was removed to remove data points with obvious errors or abnormal fluctuations, and then normalized processing was carried out to unify the data into a specific numerical interval, so as to eliminate dimensional differences and improve the stability and accuracy of subsequent model processing.

(2) secondary decomposition: First using CEEMDAN initial decomposition of data sequences, make the original complex wind speed sequence is decomposed into n with different frequency component of intrinsic mode function (IMF). Then, in view of CEEMDAN solution to get the highest frequency of IMF component, application of VMD again decomposition.

(3) Sample entropy calculation: For each subsequence obtained after CEEMDAN decomposition, the sample entropy is calculated, and the subsequence with similar sample entropy value is reorganized. After the second decomposition of VMD, the number of sequences is small, the sample entropy is small, and the recombination is not carried out.

(4) TCN network Modeling: The reconstructed signal is input into TCN for modeling.

(5) SSA hyperparameter optimization: Use SSA to optimize the hyperparameters of TCN network to get the best network structure.

(6) Forecast and results integration: The model determined by SSA optimization is used to establish a prediction model for each sub-sequence, and then the prediction results of all sub-sequences are accumulated to obtain accurate wind speed prediction results.

## 3 Model Parameter Determination

In the proposed model, some parameters need to be determined by experiments, otherwise the accuracy of the final prediction results will be adversely affected. For example, the number of decomposition layers of VMD and the network parameters of TCN affect the performance of the model. However, the number of decomposition levels of CEEMDAN is determined by the complexity of the wind speed sequence itself. The data in this paper are from the measured wind speed data of a power station in the Pacific Ocean of the United States. The power station recorded data once a day, and a total of 1,097 consecutive data samples were recorded from January 1, 2022 to December 31, 2024. Considering the beauty of the image, only the first 1000 items of data were taken as valid data, of which the first 700 measured data were the training data of the model, and the last 300 measured data were the test data of the model.

### 3.1 Evaluation Index

Scientific and reasonable quantitative comparison of the prediction results of the models can ensure that the performance of the models can be evaluated more objectively. In this paper, root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and determination coefficient R2 are selected as

evaluation indexes. The smaller the values of the first three indexes, the smaller the model prediction error, and correspondingly, the better the model performance. The closer $R^2$ is to 1, the higher the degree of fitting between the predicted value and the actual value, and the better the model performance. They are calculated as follows:

$$E_{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(p(i)-x(i))^2} \quad (24)$$

$$E_{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|p(i)-x(i)\right| \quad (25)$$

$$E_{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{p(i)-x(i)}{x(i)}\right|\times 100\% \quad (26)$$

$$R^2 = 1-\frac{\sum_{t=1}^{N}(x(i)-p(i))^2}{\sum_{t=1}^{N}(x(i)-\overline{x}(i))^2} \quad (27)$$

Formula, N number of sample points for the data; p(i) is the predicted value of the model; x(i) is the true wind speed.

### 3.2 Determination of TCN Network Parameters

The TCN parameter determines the performance of the network [33]. To effectively extract features while controlling the model size, this paper aims to investigate the optimal values for the learning rate, the number of convolutional kernels, and the regularization coefficient within the TCN parameters. Setting parameters of the search space first, and then using SSA of TCN internal super parameters optimization, select the optimal parameter combination. In the experiment, the population size of SSA algorithm was set to 8, the maximum number of iterations was set to 6, ST=0.6, PD=0.7, SD=0.2. The search space and hyperparameter optimization results of TCN parameters are shown in Table 2. Among them, the preset value range of the regularization coefficient is [1e-4,1e-1]. Since the optimal value of the regularization coefficient remains at 1E-4, it is explained separately here and not shown in the table.

**Table S2 Results of hyperparameter optimization**

| TCN parameter | Default range | Optimal value | | | |
|---|---|---|---|---|---|
| | | $IMF_1^1$ | $IMF_1^2$ | $IMF_1^3$ | $IMF_1^4$ |
| Learning rate | [1e-3,1e-2] | 0.0037 | 0.009 | 0.01 | 0.009 |
| Number of convolution nuclei | [16,128] | 64 | 16 | 16 | 64 |
| TCN parameter | Default range | $BLIMF_2$ | $BLIMF_3$ | $BLIMF_4$ | $BLIMF_5$ |
| Learning rate | [1e-3,1e-2] | 0.001 | 0.0075 | 0.01 | 0.0097 |
| Number of convolution nuclei | [16,128] | 128 | 64 | 16 | 64 |

### 3.3 VMD Decomposition Layer Number Selection

The IMF1 signal generated by CEEMDAN contains a lot of residual noise, and the frequency is at a high level, which greatly increases the difficulty of accurate prediction. To enhance the prediction accuracy, choose VMD to IMF signal for further treatment, aims to bring its subdivided into more model component, thereby more accurately capture the signal features. In order to determine the ideal number of secondary decomposition layers of VMD, the SD - SSA - TCN combined prediction model is used in this section for evaluation.

In the experimental design stage, the upper limit of decomposition layers is set to 10 layers, so as to comprehensively explore the influence of different decomposition layers on the prediction effect. Figure 10 directly presents the prediction errors of the SD - SSA - TCN model in the one-step and two-step prediction scenarios, corresponding to different VMD decomposition layers. Through careful analysis of the data in

Figure 10, it can be found that the prediction error reaches the minimum when the number of VMD decomposition layers is 4. Based on this, the number of decomposition layers of VMD is determined as 4 layers in the subsequent research work.

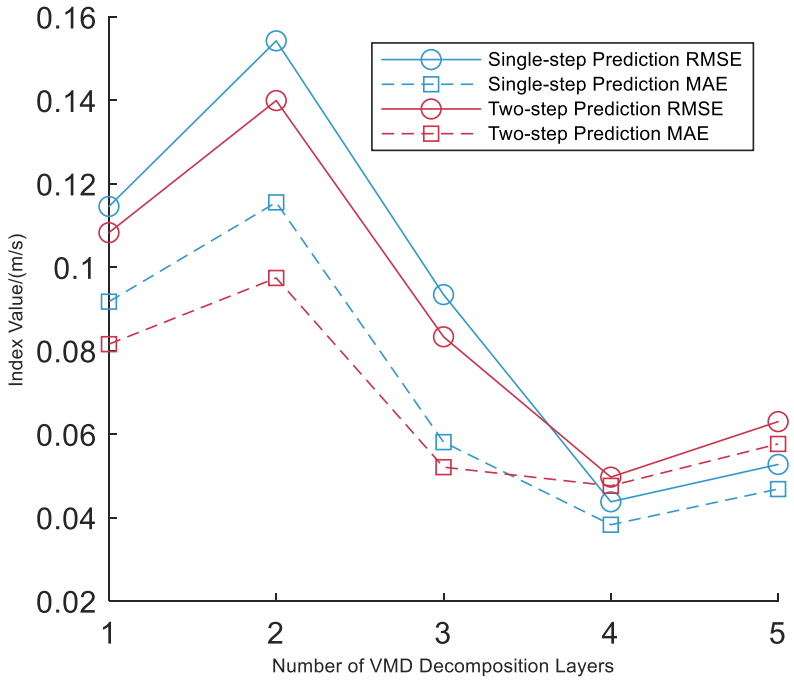



**Figure 10 $IMF_1$ prediction error after VMD decomposition**

## 4 Experimental Analysis

In this section, the SD - SE - SSA - TCN model proposed in this paper will be discussed from various performance indexes, such as model parameters, quadratic decomposition, comparison experiments, etc.

### 4.1 Model Parameter Setting

In the third section above, the optimization results of the network parameters of the TCN model are introduced, and the initial parameter setting of SSA is also explained as a prerequisite. In the comparison experiment, the initial parameter Settings of each model are shown in Table 3, and the optimizer used by all models is the Adam optimizer.

**Table S3 Parameter Settings of each model**

| Model | Argument | Numerical value |
|---|---|---|
| VMD | Penalty factor | 2000 |
| | Decomposition layer number | 4 |
| | Dc component processing | 0 |
| | Convergence threshold | 1e-7 |
| CEEMDAN | Noise standard deviation | 0.01 |
| | Number of noise implementations | 100 |
| | Maximum iterations | 100 |
| | Signal to noise ratio processing sign | 1 |
| BP | Enter the number of layer nodes | 12 |
| | Output the number of nodes in the layer | 1 |
| LSTM | Number of hidden layer neurons | 30 |

| | |
|---|---|
| Maximum iterations | 1000 |
| Gradient norm threshold | 1 |
| Learning rate | 0.0045 |

### 4.2 Validity Verification of Quadratic Decomposition

To systematically evaluate the predictive optimization capability of the dual-stage decomposition strategy, a hybrid-mode comparative experiment was designed as follows: Initially, a CEEMDAN - based decomposition-prediction benchmark model was established using raw signals, whereby feature components extracted via CEEMDAN were independently predicted through distinct prediction architectures (BP, LSTM, and SSA - TCN). Subsequently, VMD was integrated to construct a cascaded CEEMDAN - VMD decomposition framework, and collaborative component predictions were conducted using the identical model ensemble. This methodology rigorously validates the contribution of dual-stage decomposition to non-stationary feature decoupling, particularly in resolving cross-component nonlinear interactions. Considering that $IMF_1$ after CEEMDAN decomposition is the most representative and can save time and cost, $IMF_1$ in this experiment is the prediction sample, and the number of decomposition layers of VMD is set to 4 layers. The experimental results are shown in Table 4. The performance evaluation index of the model is shown in the bar chart to observe the performance of each model more directly. The single-step error is shown in Figure 11, and the multi-step error is shown in Figure 12.

**Table S4 Decomposition prediction error** **Unit: m/s**

| Decomposition model | Prediction model | Step size | RMSE | MAE |
|---|---|---|---|---|
| Primary decomposition | CEEMDAN - BP | Single step | 0.8860 | 0.6760 |
| | | Double step | 0.9504 | 0.7129 |
| | CEEMDAN - LSTM | Single step | 1.6935 | 1.59013 |
| | | Double step | 2.1045 | 1.4015 |
| | CEEMDAN - SSA -TCN | Single step | 0.8251 | 0.6044 |
| | | Double step | 0.90686 | 0.69206 |
| secondary decomposition | SD -BP | Single step | 0.0978 | 0.0748 |
| | | Double step | 0.1237 | 0.0935 |
| | SD -LSTM | Single step | 0.2276 | 0.1570 |
| | | Double step | 0.3026 | 0.2086 |
| | SD -SSA -TCN | Single step | 0.0516 | 0.0349 |
| | | Double step | 0.1112 | 0.0822 |

It can be seen from the data in Table 3 that no matter BP, LSTM or SSA -TCN are used as the benchmark model, the combined prediction model constructed with the quadratic decomposition strategy has better prediction accuracy. Taking single-step prediction as an example, RMSE decreased by 90.8%, 86.0% and 81.2% respectively in the model optimized by quadratic decomposition compared with the model only using CEEMDAN decomposition. MAE also decreased by 84.1%, 87.78% and 81.1%, respectively.

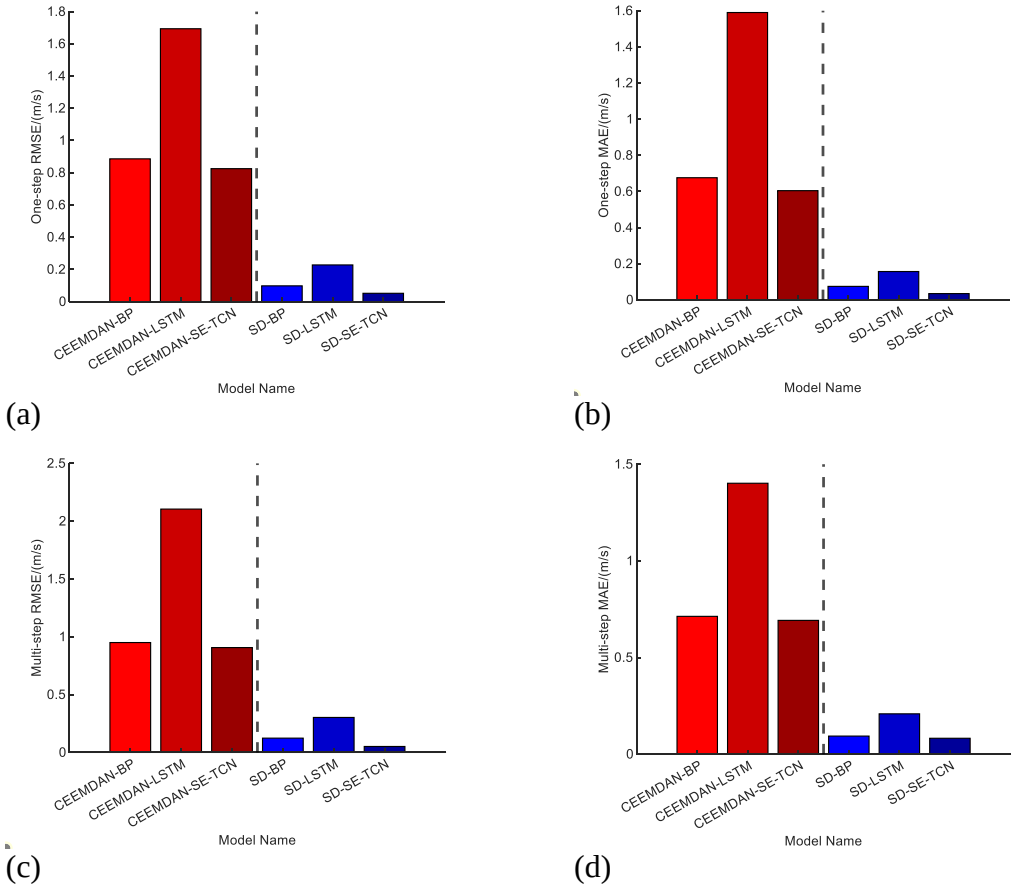



**Figure 11 Performance indicators of each decomposition model**

The index analysis in Figure 11 shows that after CEEMDAN-VMD two-stage decomposition, the mixed high-frequency noise components in the original $IMF_1$, are significantly suppressed, and these sub-sequences after secondary decomposition effectively alleviate the chaotic characteristics of the wind speed sequence. The experimental data confirm that the quadratic decomposition strategy for the dominant mode can significantly enhance the timing response accuracy of extreme wind speed events through the physical interpretability optimization of signal reconstruction.

### 4.3 Effective Verification of SSA -TCN Model

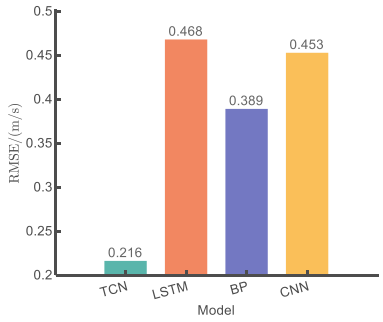
In order to comprehensively and effectively verify the validity and superiority of the proposed model, two sets of comparison experiments were carefully designed and carried out, focusing on the comparative analysis under the single step prediction scenario. In this experiment, the single-step prediction experiment will be further divided into two different experimental strategies: single model comparison experiment and mixed model comparison experiment.

In the single-model prediction experiments, a controlled benchmarking protocol was implemented to rigorously compare the proposed TCN architecture against three established reference models: CNN, LSTM, and BP neural networks. Through systematic evaluation of multi-dimensional performance metrics (including but not limited to forecasting accuracy, computational efficiency, and generalization capability), this investigation quantitatively delineates operational disparities among these models within the single-step forecasting paradigm, particularly highlighting the TCN's superior temporal feature extraction capacity as evidenced by its statistically significant outperformance in nonlinear signal approximation tasks. The performance evaluation indicators of each model are shown in Table 5, and the bar chart of performance indicators is shown in Figure 12.

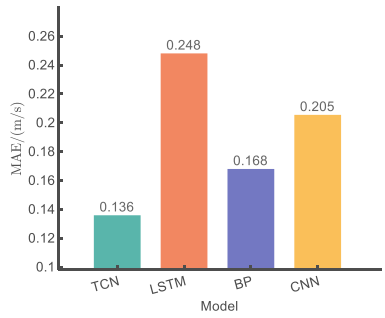
**Table S5 Prediction error of single model**

| Prediction | RMSE | MAE | MAPE | $R^2$ |
|---|---|---|---|---|

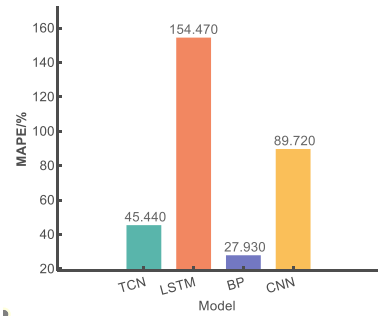
| model | | | | |
|---|---|---|---|---|
| TCN | 0.2161 | 0.1359 | 45.44% | 0.893 |
| LSTM | 0.4678 | 0.2481 | 154.47% | 0.732 |
| BP | 0.3889 | 0.1680 | 27.93% | 0.857 |
| CNN | 0.4527 | 0.2055 | 89.72% | 0.809 |



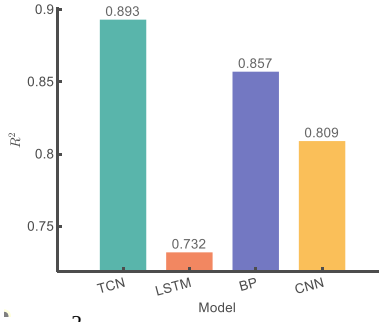

(a) RMSE




(b) MAE

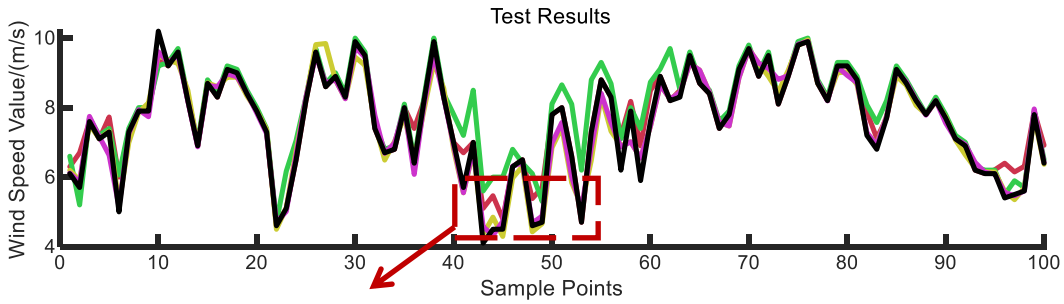



(c) MAPE

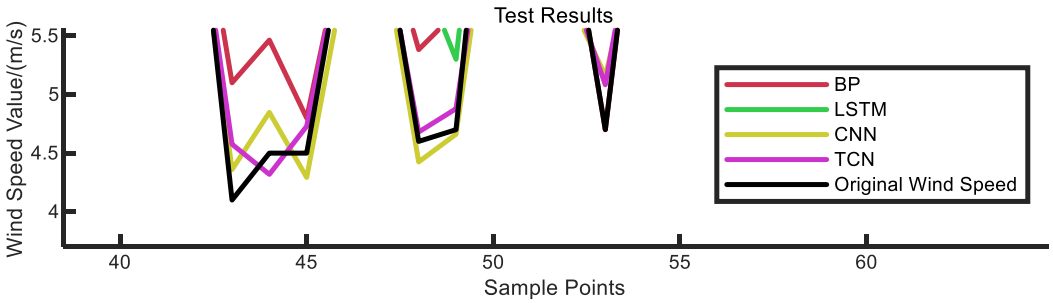



(d) $R^2$

**Figure 12 Single model indicator performance**




(a) 100 sample points prediction map




(b) Enlarge figure part sample points prediction

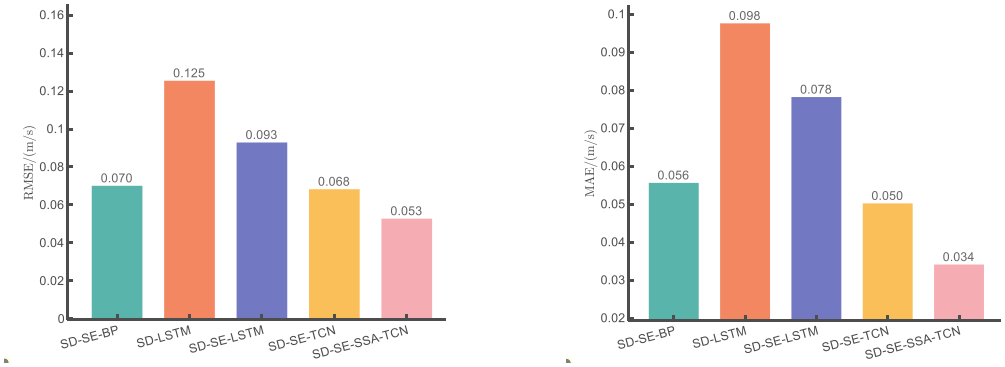
**Figure 13 Prediction results of single model test set**

When conducting mixed model comparison experiments, this paper conducted mixed model comparison experiments between SD -SE - SSA - TCN and SD -SE -BP, SD -LSTM, SD -SE - LSTM and SD -SE -TCN. The results are shown in Table 6, and the bar chart of performance

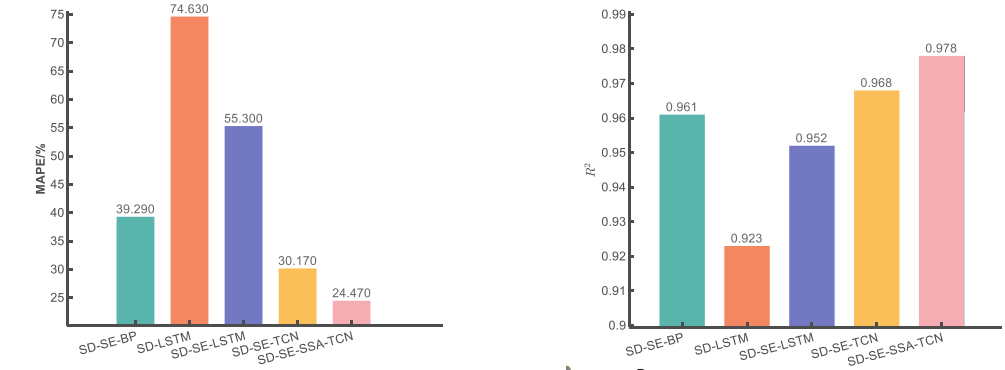
indicators is shown in Figure 14. In addition, Figures 13 and 15 respectively show the images obtained by selecting 100 predicted points in the two sets of comparative experiments, and some of the sample points are enlarged for more intuitive observation.

**Table S6 Prediction errors of mixed models**

| Prediction model | RMSE | MAE | MAPE | $R^2$ |
|---|---|---|---|---|
| SD -SE -BP | 0.0701 | 0.0556 | 39.29% | 0.961 |
| SD -LSTM | 0.1254 | 0.0976 | 74.63% | 0.923 |
| SD -SE -LSTM | 0.0929 | 0.0782 | 55.30% | 0.952 |
| SD -SE -TCN | 0.0683 | 0.0502 | 30.17% | 0.968 |
| SD -SE -SSA -TCN | 0.0528 | 0.0341 | 24.47% | 0.978 |

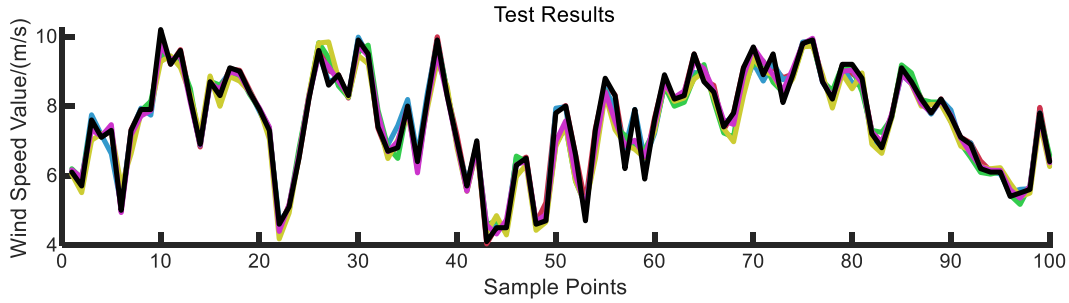



(a) RMSE (b) MAE

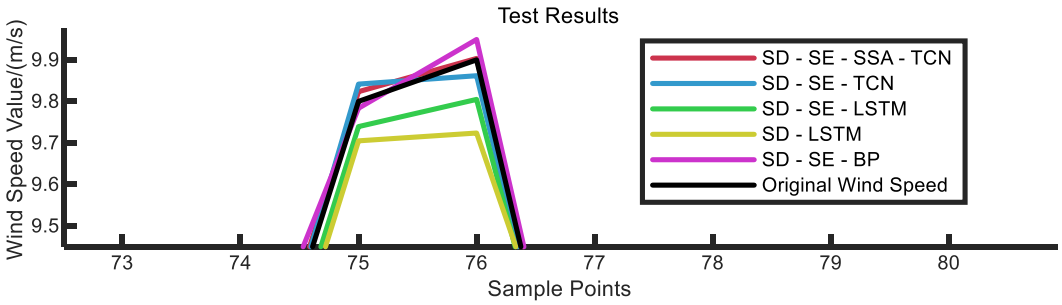



(c) MAPE (d) $R^2$

**Figure 14 Performance indicators of the mixed model**




(a) 100 sample points prediction map




(b) Enlarge figure part sample points prediction

**Figure 15 Prediction results of the mixed model test set**

### 4.4 Analysis of Experimental Results

First of all, it can be seen from Table 5 that RMSE and MAE of TCN are the best indicators, and MAPE of TCN is larger but only higher than that of BP model. This can be confirmed in Figure 13, where the wind speed prediction curve of the TCN model is more in line with the curve of the original wind speed series. BP neural network, as the most general neural network at present, shows better ability under large data samples, and has better fitting performance than LSTM and CNN. However, when LSTM is used, its hyperparameters are not optimized, and LSTM has high computational requirements on the volatility and length of data, which may be a factor leading to its unsatisfactory effect. In summary, the TCN model, which incorporates extended convolution technology, convolutional layers, and parameter sharing, demonstrates superior performance compared to other benchmark models.

Secondly, the performance of the single model is compared with that of the corresponding mixed model. When it has been proved that the quadratic decomposition is better than the single CEEMDAN decomposition, the mixed model adopts the quadratic decomposition. Contrast table 5 and table 6 calculates the RMSE of SD - SE - LSTM model, MAE, MAPE LSTM compared with single model improved by 80.1%, 68.4% and 64.2%, respectively, RMSE, MAE and MAPE of SD -SE -SSA -TCN model were improved by 75.6%, 74.9% and 46.1%, respectively, compared with single model TCN. This shows that compared with using a single model to predict the original data directly, proper preprocessing can effectively improve the prediction performance of the model.

Finally, Figure 15 intuitively illustrates that the SD -SE -SSA -TCN model proposed in this paper is closest to the original wind speed sequence and has the best fitting effect compared with the other four hybrid models. The four indicators of this model are also better than other models, and the coefficient of determination reaches 0.978, which is close to 1. At the same time, from the comparison of SD -LSTM and SD -SE -LSTM results in Table 6, that is, the recombinant sequence prediction error MRSE and MAE decreased respectively, it can be seen that whether sample entropy recombination has a great impact on the accuracy of the result prediction. The subsequence reorganized by sample entropy improves the generalization ability of the model and reduces the fluctuation of data. In addition, the SSA algorithm can avoid the model falling into the local optimal solution. It is found that the SD -SE -SSA -TCN model can effectively avoid the deviation caused by manual screening of network hyperparameters, and effectively improve the overall performance of the network. Compared with SD -SE -TCN, MRSE, MAE and MAPE of SD -SE -SSA -TCN were reduced by 22.7%, 32.1% and 18.9%, respectively. In general, quadratic decomposition, sample entropy recombination and sparrow search algorithm have improved the prediction effect of the network. The best performance can be obtained by combining the quadratic decomposition, SE and SSA -TCN. Taking SD -SE -BP and SD -SE -LSTM as examples, RMSE of SD -SE -SSA -TCN decreased by 24.6% and 43.2% in the forecast, MAE decreased by 38.7% and 56.4%, MAPE decreased by 37.7% and 55.8%, and R2 increased by 1.73% and 2.66%.

### 5 Summary

This study addresses the dual challenges of non-stationary wind speed sequences and the limited adaptability of conventional forecasting models

by proposing a hybrid forecasting framework integrating dual-stage decomposition (CEEMDAN -VMD), sample entropy reconstruction, and Sparrow Search Algorithm-optimized temporal convolutional networks (SSA -TCN). The proposed architecture systematically implements a triple-coupled "decomposition-entropy reconstruction-parameter self-optimization" mechanism, thereby demonstrably enhancing forecasting precision for non-stationary wind speed series through coordinated multi-scale feature decoupling and adaptive neural topology configuration.

(1) CEEMDAN was used to decompose the original wind speed sequence for the first time, and the sample entropy was used to carry out adaptive recombination of the high-frequency IMF component. Combined with the secondary decomposition of VMD, the random noise of the sequence was reduced (the sample entropy of IMF1 component was reduced by 85.03%), and the data stability was significantly improved.

(2) Construct a deep TCN network framework, introduce time series attention mechanism to enhance the extraction ability of historical wind speed multi-scale features, and design Sparrow algorithm to dynamically optimize TCN learning rate, convolution kernel size and other hyperparameters to avoid the local optimal problem of manual parameter adjustment.

(3) Experimental results of measured wind speed series verify the effectiveness of the collaborative work of quadratic decomposition, entropy recombination and intelligent optimization proposed in this paper. Compared with SD -LSTM, SD -SE -LSTM, TCN, SD -SE -TCN and other models, RMSE, MAE, MAPE and other indicators of prediction errors of the proposed method are significantly reduced.

In this paper, the step size of TCN is determined according to experience, and it cannot be determined adaptively by more scientific and effective optimization algorithm. At the same time, the exponential growth characteristic of TCN expansion coefficient is not suitable for multi-scale wind speed characteristics. In future studies, we will further explore how to determine the structural parameters of expansion convolution adaptively, and introduce physical constraint equations to enhance the robustness of prediction of extreme wind speed events, so as to make collaborative prediction of the volatility of wind farm clusters.

**Statement of Ethics**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Conflict of Interest Statement**

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

**Funding Sources**

Wuhan University of Science and Technology, Application of bidirectional synchrosqueezing transform in weak fault detection of bearing cage, Z202302

**References**

1. Lu Yi Feng, Wang Xiao. Short-term wind power prediction model based on quadratic decomposition and IDBO-DABiLSTM [J]. Computer engineering, 2024, 50 (12) : 99-109. The DOI: 10.19678 / j.i SSN. 1000-3428. 006 8365.
2. Yin Hao, Ou Zuhong, Chen De, et al. Ultra-short term wind power prediction based on quadratic mode decomposition and Cascading deep learning [J]. Power Grid Technology, 2020,44(02):445-453.DOI:10.13335/j.1000-3673.pst.2019.1266.
3. Xiang Ling, Liu Jianing, Su Hao, et al. Multi-step wind speed prediction based on CEEMDAN quadratic decomposition and LSTM [J]. Journal of Solar Energy,2022, 43(08):334-339.DOI:10.19912/j.0254-0096.tynxb.2020-1410.
4. Wang Shengyan, Wang Juanjuan. Short-term wind speed prediction based on CEEMDAN-SE-GGO-LSTM model [J]. Electrician

technology, 2024, (4) : 74-78 + 81. DOI: 10.19768 / j.carol carroll nki DGJS. 2024.04. 018.
5. Yin Yuan-Ya, Pan Wen-Hu, ZHAO Wen-guang, et al. Ultra-short term Wind speed prediction method based on CEEMDAN and BiLSTM-AM [J]. Electrical testing and instrument, 2024, 21 (9) : 77-84. The DOI: 10.19753 / j.i ssn1001-1390.2024.09.010.
6. Gao Shengyang, Li Fashe. Short-term wind speed prediction based on CEEMD-SE-PSR-BP [J/OL]. Journal of solar energy, 1-7 [2025-03-11]. https://doi.org/10.19912/j.0254-0096. tynxb.2023-2035.
7. Tang Fei, Li Hao. Optimization of wind speed prediction model of echo state Network based on variational mode decomposition and whale algorithm [J]. Journal of Sensing Technology, 2018,37(10):1770-1777.
8. Wang Jun, Li Xia, Zhou Xidong, et al. Based on super short term wind speed forecasting of VMD and LSTM [J]. Power system protection and control, 2020 (11) : 13 45 to 52. DOI: 10.19783 / j.carol carroll nki PSPC. 190860.
9. Li Chunmei, Gu Jiacheng, Wang Shanqiu, et al. Optimization of short-term wind speed prediction for GRU based on improved Marine Predator algorithm [J]. Journal of Baicheng Normal University, 2019,38(05):29-39. (in Chinese)
10. Shang Liqun, Li Hongbo, Hou Yadong, et al. Short-term PV power prediction based on VMD-ISA-KELM [J]. Power system protection and control, 2022, 50 (21) : 138-148. The DOI: 10.19783 / j.carol carroll nki PSPC. 220140.
11. Gong H ,Xing H ,Wang Q .Enhanced forecasting method for realized volatility of energy futures prices: A secondary decomposition-based deep learning model[J].Engineering Applications of Artifi - cial Intelligence, 2025146103, 21-1103 21.
12. ZHAO Zheng, Nan Honggang, Qiao Jintao. Research on ultra-short term Wind speed prediction based on improved time series based on quadratic decomposition [J]. Journal of North China Electric Power University (Natural Science Edition),2020,47(04):53-60.
13. Tang F .Short-term wind power prediction based on improved sparrow search algorithm optimized long short-term memory with peephole connections[J].Wind Engineering, 2019,49(1):71-90.
14. Naheliya B ,Kumar K ,Redhu P .A deep extreme learning machine approach optimized by sparrow search algorithm for forecasting of traffic flow[J].Physica Scripta, 2019,99 (12):125288-125288. (in Chinese)
15. Zhou D ,Liu Y ,Wang X , et al.Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM - TCN - BiLSTM [J]. Energy, 47, 2025318348-134847.
16. Wei Pengfei, Fan Xiaochao, Shi Ruijing, et al. Short-term photovoltaic power prediction based on improved Sparrow search algorithm and optimized support vector machine [J]. Thermal power generation, 50, 2021 (12) : 74-79. The DOI: 10.19666 / j.r LFD. 202104127.
17. Fu Yang, Ren Zixu, Wei Shurong, et al. Ultra-short term power prediction of offshore wind power based on improved LSTM-TCN model [J]. Proceedings of the CSEE,2022,42 (12): 4292-4303. (in Chinese) DOI:10.13334/ j.025 8-8013.pcsee.210724.
18. Jiankai Xue, Bo Shen. A novel swarm intelligence optimization approach: sparrow search algorithm[J]. Systems Science & Control Engineering, 2019,8(1): 22-34.
19. Zhao Lingyun, Liu Youbo, Shen Xiaodong, et al. Short-term wind power prediction model based on CEEMDAN and improved time Convolutional network [J]. Protection and control of electric power systems, 2022, 50 (01) : 42-50. DOI: 10.19783 / j.carol carroll nki PSPC. 210252.
20. Li Feihong, Xiao Yingqun. Based on EMD TCN - short-term power load forecasting of ELM [J]. Computer system application, 2022, 31 (11) : 223-229. The DOI: 10.15888 / j.carol carroll nki. Csa. 008781.