

A Wind Energy Prediction Scheme Combining Cauchy Variation and Reverse Learning Strategy

Xiaokun WU¹, Xiaoyu SHEN¹, Jinghui ZHANG¹, Yagang ZHANG^{1,2}

¹*Institute of Data Science and Statistical Analysis, North China Electric Power University, Baoding, 071003, China*

²*Interdisciplinary Mathematics Institute, University of South Carolina, Columbia, SC 29208, United States*
yagangzhang@ncepu.edu.cn

Abstract—Based on climate change, the mission of a major country, energy security, and low-carbon economy, carbon emission reduction has been raised to a global strategic level, and the world's energy transformation has a long way to go. Based on the characteristics of random volatility and intermittency of wind power itself, this paper focuses on improving the accuracy of forecasting, analyzes the complex new energy forecasting problems, and proposes the following hybrid forecasting scheme: First, high frequency denoising of wind power sequence using synchrosqueezed wavelet transforms (SWT), reducing the influence of noise on prediction performance, and improving time-frequency aggregation. Then use the long short-term memory model (LSTM) with long-term information memory ability to predict the denoising sequence, and effectively extract the non-linear sequence mapping information. In the meantime, the sparrow search algorithm (SSA) that combines cauchy mutation and opposition-based learning strategy is added to optimize the selection of model hyperparameters, further enhance the reliability of model prediction, and alleviate the burden of model training. Verified by different wind farm cases, the proposed hybrid forecasting model has obvious performance advantages, is conducive to the safe and high-reliable of the power system.

Index Terms—carbon emissions, Cauchy mutation, long short-term memory, reverse learning, synchrosqueezed wavelet transforms.

I. INTRODUCTION

The form of energy consumption affects the economic development to a great extent, and the climate change caused by energy consumption has also become a global problem faced by human beings. Against this backdrop, countries have agreed to reduce greenhouse gas emissions, and many have elevated the goal to the status of national strategy, which means that the international energy mix urgently needs to develop an efficient energy system based on renewable energy. For this reason, clean energy, led by wind energy, is at the heart of the energy transition. However, the wind power system has many problems of its own, including random volatility, intermittent, poor robustness in response to extreme events, which will bring high uncertainty risk to the supply side and the demand side of the new energy power system, thus affecting the capacity

of the grid, and bringing great threat to the stable and safe operation of the grid. In order to get rid of this dilemma, it is very necessary to predict wind power with high precision and to accurately grasp its change rule in a certain period of time in the future.

According to the results of the present study can forecast wind power related technical summary for the following four categories: based on physical information, mining information by statistical methods, the application of intelligent algorithms, and combined models. At present, most physical modeling methods are NWP assessment of wind energy resources [1-2]. Although the relationship between physical information and prediction object is fully considered, the calculation is complex. Then, the statistical method of establishing the mapping relationship between historical data and prediction objectives is widely used in wind energy evaluation [3-4], but it is difficult to make effective prediction in the absence of a large number of historical data. Therefore, artificial neural network as a typical intelligent technology has become one of the important methods in the forecast field [5-6]. However, there are some shortcomings, such as poor interpretability of results and too long training time. Moreover, single model cannot fully acquire data information and make effective predictions. So that hybrid methods, which combine the strengths of different forecasting methods, are popular [7].

In order to improve the learning ability of prediction models and further improve the accuracy of wind speed prediction, traditional algorithms including genetic algorithm [8], cuckoo algorithm [9], Multi-parametric Quadratic Programming [10] are widely used in parameter optimization. At the same time, some new algorithms: Slime Mould Algorithm [11], Conflict monitoring optimization heuristic [12] have also been proposed. However, according to the theory of no free lunch, no algorithm is optimal for all problems, so it is possible to obtain a good optimization effect by choosing a suitable optimization algorithm for a specific problem.

In this article, a new hybrid wind speed forecast framework is presented, which integrates data denoising and improved prediction algorithm, which reduces the influence of noise on the sequence prediction, while improving the robustness and anti-interference ability of the model. The contributions of this article are described below:

(1) Synchrosqueezed wavelet transforms is adopted to

This research was supported partly by the National Natural Science Foundation of China (51637005), the S&T Program of Hebei (G2020502001), and the Fundamental Research Funds for the Central Universities (2018MS127).

eliminate high frequency noise points and improve energy aggregation on the premise of perfect signal reconstruction, which reduces the negative impact of random noise in wind power sequence on subsequent prediction work.

(2) Compared with other traditional neural networks, LSTM has the ability to learn long-term dependent information, which can deeply explore the complex mapping relationship between variables, and effectively reflect the characteristics of data fluctuations as well. Therefore, this model is selected for prediction.

(3) In order to set the hyperparameters of the neural network, most studies adopt manual parameter adjustment according to experience, which is not only inefficient, but also may avoid the optimal parameter region. Based on this, the application of sparrow search algorithm (SSA) in LSTM network parameter adjustment is beneficial to raise the effectiveness of prediction and decrease the training cost of neural network. In order to avoid the SSA sinking into the local optimal dilemma, cauchy mutation and back learning strategies are introduced to further raise the performance of the mixed model.

The design framework of the remaining part of this study, including synchrosqueezed wavelet transforms and other methods involved in the paper will be introduced in detail in the second part, the modeling process of the paper is listed in detail in the third part, the fourth part is case analysis and model evaluation, and the final conclusion is presented in the fifth part.

II. BASIC ALGORITHM OF ESTABLISHING MODEL

A. Synchrosqueezed Wavelet Transforms

Synchrosqueezed wavelet transforms (SWT) was presented in 2011 [13]. It aims to further understand the role of empirical mode decomposition (EMD) and provide new ideas. Although it contains the principle of empirical mode decomposition (EMD), they are different in reconstruction. When performing time-frequency analysis, the synchrosqueezed wavelet transforms requires that the original signal be subjected to continuous wavelet transform first, and the corresponding wavelet coefficients are compressed and reordered in the frequency direction, so as to avoid the time-frequency ambiguity caused by the wavelet transform, achieve the purpose of high time-frequency resolution and further enhance the filtering and denoising effect.

Since the synchrosqueezed wavelet transform was proposed, it is often used to extract and reconstruct signals due to its strong anti-interference and robustness. The basic principles and steps are described below:

(a) Suppose any signal is formed by a series of harmonic superposition, denoted as $y(t)$. After continuous wavelet transform (CWT), the corresponding wavelet transform coefficient spectrum is $W_y(a, b)$:

$$W_y(a, b) = \int y(t) a^{-1/2} \overline{\psi\left(\frac{t-b}{a}\right)} dt \quad (1)$$

In the above formula, a , b respectively represent the expansion scale and phase variables, $\psi(\bullet)$ represents the wavelet basis function, and the corresponding $\overline{\psi(\bullet)}$ is its conjugate function. Based on Plancherel theory, the above

formula can also be equivalent to formula (2):

$$W_y(a, b) = \frac{1}{2\pi} \int \hat{y}(\xi) a^{1/2} \overline{\hat{\psi}(a\xi)} e^{ib\xi} d\xi \quad (2)$$

Among them, $\hat{y}(\xi)$ is the result of Fourier transform of the original signal, and ξ represents the angular frequency.

(b) Theoretically, the continuous wavelet transform coefficient will $W_y(a, b)$ focus on the corresponding center frequency line when a certain value ξ is obtained, but the fact is that it will always spread to the local area of the frequency line, resulting in time-frequency blur. Nevertheless, the phase variable b still contains the frequency $y(t)$. Therefore, in the case of any a , b of which have $W_y(a, b) \neq 0$, the instantaneous frequency $\omega_y(a, b)$ can be obtained by deriving $W_y(a, b)$ and the mapping relationship can be established:

$$\omega_y(a, b) = -i(W_y(a, b))^{-1} \frac{\partial}{\partial b} W_y(a, b) \quad (3)$$

(c) According to the mapping relationship established in the previous step, the synchrosqueezed wavelet transform is performed. That is, the conversion process of $W_y(a, b)$ from the time-scale surface to the time-frequency surface is completed, where (b, a) represents the time-scale plane and $(b, \omega_y(a, b))$ represents the time-frequency surface. Further, in order to avoid blurring of the time-frequency line, the wavelet change coefficients in the local area of $[\omega_\ell - \frac{1}{2}\Delta\omega, \omega_\ell + \frac{1}{2}\Delta\omega]$, $\Delta\omega = \omega_\ell - \omega_{\ell-1}$ of the frequency ω_ℓ are compressed onto the ω_ℓ line, thereby obtaining accurate and reliable synchronous compression transformation values:

$$T_y(\omega_\ell, b) = (\Delta\omega)^{-1} \sum_{a_k: |\omega(a_k, b) - \omega_\ell| \leq \Delta\omega/2} W_y(a_k, b) a_k^{-3/2} (\Delta a)_k \quad (4)$$

In the upper equation, k refers to the number of scaling scales.

(d) According to the following formula:

$$\int_0^\infty W_s(a, b) a^{-3/2} da = \int_0^\infty \overline{\hat{\psi}(\xi)} \frac{d\xi}{\xi} \cdot \frac{1}{2\pi} \int_0^\infty \hat{s}(\zeta) e^{ib\zeta} d\zeta \quad (5)$$

It can be seen that synchronous compression can realize the inverse transformation, that is, the raw signal can be reconfigured after the synchrosqueezed wavelet transform $y(t)$:

$$y(t) = \Re[C_\psi^{-1} \sum_\ell T_y(\omega_\ell, b)(\Delta\omega)] \quad (6)$$

which set $C_\psi = \frac{1}{2} \int_0^\infty \overline{\hat{\psi}(\xi)} \frac{d\xi}{\xi}$.

B. Sparrow Search Algorithm

Suppose there are N sparrows, the sparrow group with dimension d and its corresponding fitness value are expressed as follows:

$$X = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots \\ x_N^1 & x_N^2 & \cdots & x_N^d \end{bmatrix}, F(X) = \begin{bmatrix} f(x_1^1, x_1^2, \dots, x_1^d) \\ f(x_2^1, x_2^2, \dots, x_2^d) \\ \vdots \\ f(x_N^1, x_N^2, \dots, x_N^d) \end{bmatrix} \quad (7)$$

When foraging, individuals with better fitness values will act as discoverers to give priority to food resources and

provide foraging areas and directions for the entire group. In case of danger, judge whether to bring the group to other safe areas according to the magnitude of the danger value. The location of the discoverer is updated as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot iter_{max}}\right), R_2 < ST \\ X_{i,j}^t + Q \cdot L, R_2 \geq ST \end{cases}, j = 1, 2, \dots, d \quad (8)$$

In the formula, $X_{i,j}^t$ refers to the position of the i -th sparrow in the j -th dimension during the t -th iteration, $iter_{max}$ is the maximum number of iterations, the random numbers $\alpha \in (0,1)$ and Q obey the standard normal distribution, and L is the $1 \times d$ -dimensional identity matrix. By comparing the warning value R_2 and the safety value ST to determine whether to transfer to the safe area.

The remaining individuals except the discoverer are joiners. It will track the discoverer or compete with them for food resources during the foraging process. The location update is as follows:

$$X_p^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i^2}\right), i > \frac{N}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+ \cdot L, i \leq \frac{N}{2} \end{cases} \quad (9)$$

Among them, X_p represents the optimal position of the current discoverer, X_{worst} represents the global worst position information, $A^+ = A^T(AA^T)^{-1}$, the value of matrix A is +1 or -1, the value is random, and the dimension is $1 \times d$. It should be noted that the discoverers and joiners in the group are dynamically changing, but their proportion of the group number is constant.

Let SD denote randomly selected "investigators", which is approximately the number 10% ~ 20% in the group. Once a danger is found, remind the entire group to flee the current area. The location update formula is:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t|, f_i > f_g \\ X_{i,j}^t + k \cdot \frac{|X_{i,j}^t - X_{worst}^t|}{f_i - f_w + \varepsilon}, f_i = f_g \end{cases} \quad (10)$$

In the above formula, X_{best} and f_g respectively represent the global optimal location and fitness value, f_w represents the global worst fitness value, f_i represents the current fitness value, β and k are random numbers, the former obeys the standard normal distribution, the latter obeys the uniform distribution in the range of $[-1,1]$, and ε is a constant to prevent the denominator from being 0.

C. Improved Sparrow Search Algorithm

1) Improve the Initialization Process

The chaotic mapping relationship is introduced to modify the standard sparrow search algorithm to avoid entering into the local optimum and accelerate convergence rate of the algorithm. The following formula indicates that the Sin chaotic one-dimensional mapping model is used for population initialization. Compared with the Tent and Logistic chaotic models, the Sin model has an unlimited number of mapping folds and has better chaotic characteristics:

$$\begin{cases} x_{n+1} = \sin \frac{2}{x_n}, n = 0, 1, \dots, N \\ -1 \leq x_n \leq 1, x_n \neq 0 \end{cases} \quad (11)$$

In order to avoid the fixed point and zero point in $[-1,1]$,

$x_n \neq 0$ is needed.

2) Improve the Finder Location Update Process

On the basis of the original discoverer update formula, by adding the best solution of the former generation, the current discoverer's position is affected by both the previous discoverer and the best solution of the former generation. This improves the search accuracy to a certain extent and avoids local optimization. In the meantime, the idea of inertial weight is introduced into the above formula, so that a larger value can be used for global search in the early stage of the iteration. As the iteration increases, the dynamic weight factor automatically decreases, which is more conducive to the later local search. The improved finder location update formula is described as follows:

$$\omega = \frac{e^{2(1-t/iter_{max})} - e^{-2(1-t/iter_{max})}}{e^{2(1-t/iter_{max})} + e^{-2(1-t/iter_{max})}} \quad (12)$$

$$X_{i,j}^{t+1} = \begin{cases} (X_{i,j}^t + \omega(f_{j,g}^t - X_{i,j}^t)) \cdot rand, R_2 < ST \\ X_{i,j}^t + Q, R_2 \geq ST \end{cases}$$

Among them, ω is the dynamic weight factor, and $f_{j,g}^t$ is the global optimal value of the j -th dimension of the previous generation.

3) Improve the Location Update Process of Scout Sparrow

Through the following formula, the individual sparrow either does not swing between the best location and the poorest location, or is in a random area between itself and the optimal position:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \cdot |X_{i,j}^t - X_{best}^t|, f_i \neq f_g \\ X_{i,j}^t + \beta \cdot |X_{i,j}^t - X_{worst}^t|, f_i = f_g \end{cases} \quad (13)$$

4) Incorporating Cauchy Variation and Reverse Learning strategies

The perturbation ability based on cauchy mutation is strong, and it is introduced into the update process of the target position, which is beneficial to boost the global search capacity [14]. The following formula represents the position update method after introducing cauchy mutation:

$$X_{i,j}^{t+1} = X_{best}^t(t) + \text{cauchy}(0,1) \oplus X_{best}^t(t) \quad (14)$$

Among them, $\text{cauchy}(0,1)$ represents the standard cauchy distribution.

For the reverse learning strategy, it searches for the optimal solution through the reverse learning mechanism and evaluation mechanism, and incorporates it into the sparrow algorithm, and the following description can be obtained:

$$\begin{aligned} X_{best}^{t+1} &= ub + r \oplus (lb - X_{best}^t(t)) \\ X_{i,j}^{t+1} &= X_{best}^{t+1}(t) + b_1 \oplus (X_{best}^t(t) - X_{i,j}^{t+1}(t)) \\ b_1 &= (iter_{max} - \frac{1}{iter_{max}}) \end{aligned} \quad (15)$$

In the formula, lb , ub represents the upper and lower bounds, the random matrix r with dimension $1 \times d$ obeys the uniform distribution in $(0,1)$, and b_1 is the information exchange control parameter.

Although the search area of the operator can be expanded based on the above, so as to further enhance the algorithm's optimization capacity, the probability P_s of the following formula (16) is selected to determine the alternate use of reverse learning and cauchy mutation, if $rand > P_s$ is selected cauchy variation solution, on the contrary choose back

learning strategy. Note that the θ in the above formula is an adjustable parameter, generally 0.05.

$$P_s = -\exp\left(1 - \frac{1}{iter_{max}}\right)^{20} + \theta \quad (16)$$

The greedy rule is introduced to determine whether the new position obtained by the disturbance mutation strategy is better than the fitness value of the old position. The following formula (17) describes the greedy rule:

$$X_{best} = \begin{cases} X_{i,j}^{t+1}, f(X_{i,j}^{t+1}) < f(X_{best}) \\ X_{best}, f(X_{i,j}^{t+1}) \geq f(X_{best}) \end{cases} \quad (17)$$

The flow of the sparrow search algorithm based on the above improvement (ISSA) is shown as following:

(a) Initialize the parameters first: the number of the entire group, the maximum iteration times, the proportion of discoverers, the proportion of scouts and so on. Then use the Sin chaotic mapping model to initialize the population position;

(b) Calculate the adaptive value corresponding to each sparrow in the group, and find the current optimal and poorest adaptive values, as well as the corresponding positions;

(c) Select the discoverer from the top ranked sparrows in fitness, and modify the position of the discoverer;

(d) Based on the subscribers obtained in step (c), use the subscriber location update formula to make changes;

(e) Randomly choose some individuals from the sparrow group as scouts, and update their positions based on the detection and warning behavior;

(f) Choose the reverse learning strategy or the Cauchy mutation method to generate new solutions based on the probability P_s , and determine whether to update the position based on the greedy rule;

(g) Determine whether the process is terminated according to the termination condition, if so, then output the optimal solution, otherwise return to step (b).

III. MODEL ITERATION AND EVALUATION

In this paper, firstly, the SWT is used to denoise the data, and the LSTM model with long-term memory is selected for prediction [15]. At the same time, the SSA combined with Cauchy variational and reverse learning strategy is introduced to optimize the model hyperparameters to further improve the forecast performance.

The steps of the mixed model established in this research are as following:

Step 1. Data Selection

In an effort to prove the feasibility of the method described in this article, the data test models of different wind farms are selected [16]. The training set and test set are divided according to a certain proportion. The training data set is used for the model learning process, and the test set is used for evaluation.

Step 2. Data Processing

Considering that the wind speed has significant uncertainty, choosing appropriate methods to reduce the noise of the original data can better achieve the optimal training effect and make the following work better, which is difficult to achieve by using the original data without

processing modeling [17-18]. In this paper, the results of continuous wavelet transform are squeezed to reduce the random noise, and then the original signal is recovered by inverse transform, so as to achieve the purpose of suppressing the noise.

Step 3. Model Selection

In this paper, LSTM model with long short-term memory function is used to predict wind speed because of the randomness and uncertainty of wind speed [19-21]. This model can combine the effects of the long-ago state and the recently concluded state, and overcome the problems of gradient disappearance and eruption of traditional neural network.

Step 4. Model optimization

Deep learning hyperparameter setting generally adopts empirical method without clear guidance, which often leads to a large amount of time cost to try the parameters with the best effect [22-24]. Therefore, ISSA is used to optimize the hidden layer nodes, training times and learning rate of LSTM, and Mean absolute error is selected as the algorithm fitness function to improve the optimization efficiency.

Step 5. Model Evaluation

Different models are used for prediction, and the proposed method is compared and evaluated to test its effectiveness [25]. Mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) were selected for error analysis. The calculation formula is as below:

$$MAE = \frac{1}{n} \sum_{t=1}^n |z_t - \hat{z}_t| \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (z_t - \hat{z}_t)^2} \quad (19)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{z_t - \hat{z}_t}{z_t} \right| \quad (20)$$

where, z_t is the observed value, \hat{z}_t is the result, n is the number of sequences.

In this paper, synchrosqueezed wavelet transforms are used for noise reduction, and LSTM is selected to establish the wind speed prediction model. Considering the influence of parameters on the results in the learning process, ISSA is used for optimization to improve the learning performance and generalization ability of neural network model and avoid over fitting.

The modeling process is displayed in Figure 1.

IV. CASE ANALYSIS AND EVALUATION

Spanish wind farm: Sotavento with an interval of 10 minutes and Chinese wind farm: Changma with an interval of 5 minutes are selected as samples to compare the effect of this model. 1000 samples were selected, of which 700 were used for model learning and the rest for verification. In Figure 2, (a) and (b) are two groups of wind speed data respectively. As shown in the figure, the 700 train data of the two wind farms are represented by blue lines, while the remaining 300 test data are represented by red lines. In order to clearly indicate the grouping boundary of the two types of data, the black dashed line is used to group the 700th point to area segmentation.

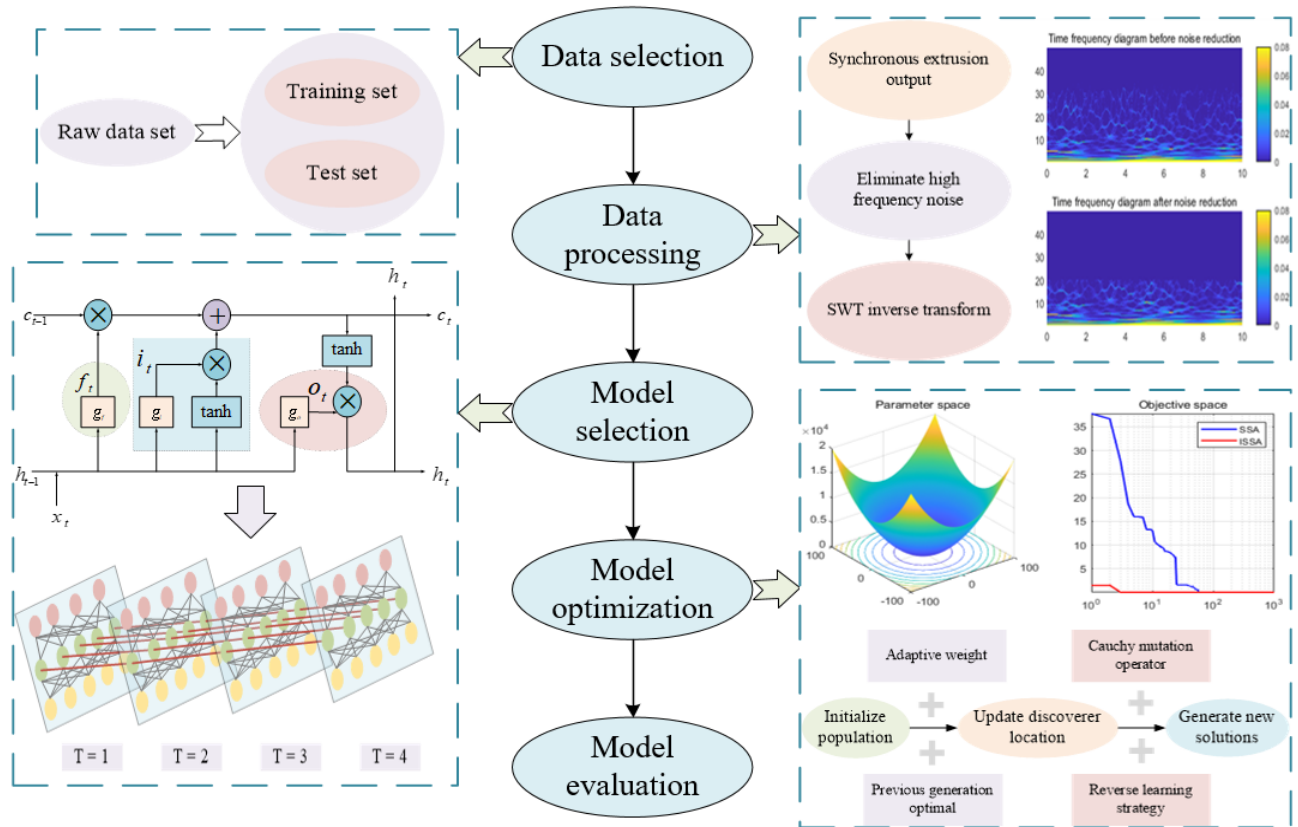


Figure 1. Flow chart of model establishment

Both groups of samples have obvious nonlinear characteristics and volatility, as well as insignificant periodicity, as the peaks and troughs of the two sets of data have different positions, and the data trend is obviously different. The comparative test using these two groups of data can effectively show that the scheme proposed in this paper has good performance in different wind speed data, and can provide support for wind power dispatching.

Table 1 lists the basic statistical characteristics of data. According to the data in the table, compared with Sotavento wind farm, the maximum value of Changma wind farm data is 18.91m/s, which is significantly higher and the degree of dispersion is greater. Combined with the figure, Sotavento wind farm fluctuates in the short term, but the overall variance is small; the data of Changma wind farm has greater fluctuation, obvious mutation characteristics, and stronger randomness and uncertainty.

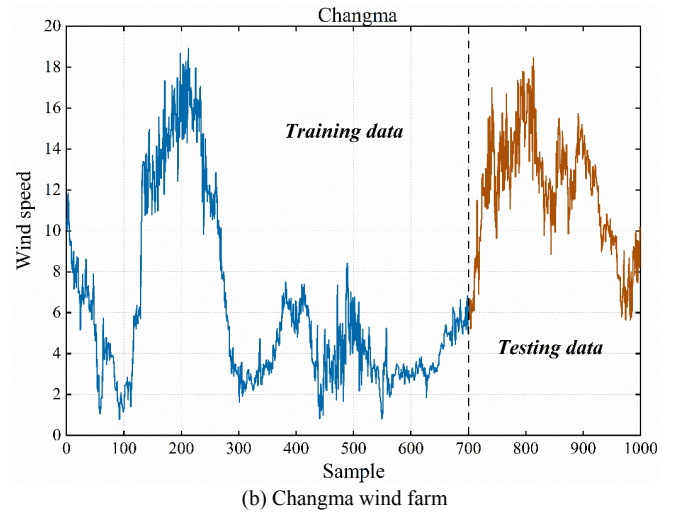
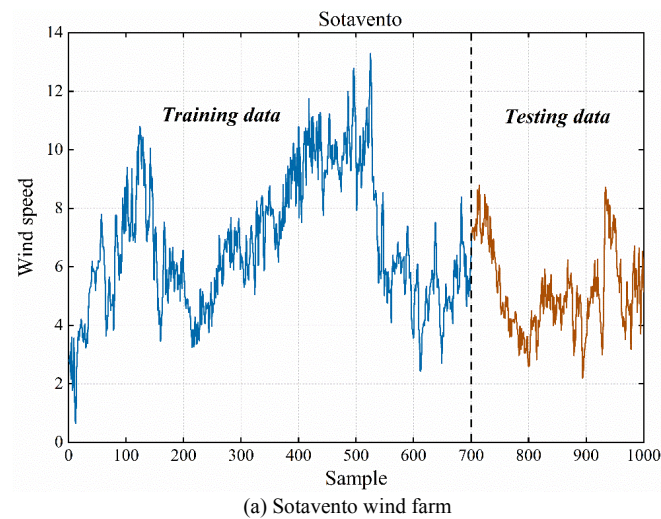


Figure 2. Raw data sequence

TABLE I. BASIC STATISTICAL CHARACTERISTICS OF WIND SPEED DATA

Wind Farm	Max(m/s)	Min(m/s)	Mean(m/s)	Std(m/s)
Sotavento	13.29	0.65	6.28	2.16
Changma	18.91	0.79	7.88	4.71

By introducing a denoising algorithm, SWT, you can view changes in data fluctuation noise and the magnitude and direction of changes in model prediction performance. Then the SSA intelligent algorithm is used to optimize the parameters of the selected model LSTM to improve the performance of the model, and for the limitation of algorithm convergence, the improved SSA algorithm is used to optimize the selected model and determine the final hybrid forecasting model framework. At the same time, in order to further verify the feasibility of the proposed hybrid model, based on the error indicators MAE, RMSE and

MAPE, the proposed method is compared with the single model (LSTM) and the hybrid model (SWT-LSTM, SWT-SSA-LSTM). Based on two sets of wind farm data at different time intervals, a total of eight sets of detection experiments are presented to highlight the importance of data denoising processing and model optimization in wind power forecasting. In addition, it should be noted that the parameter settings related to the same model are consistent.

A. Data Denoising

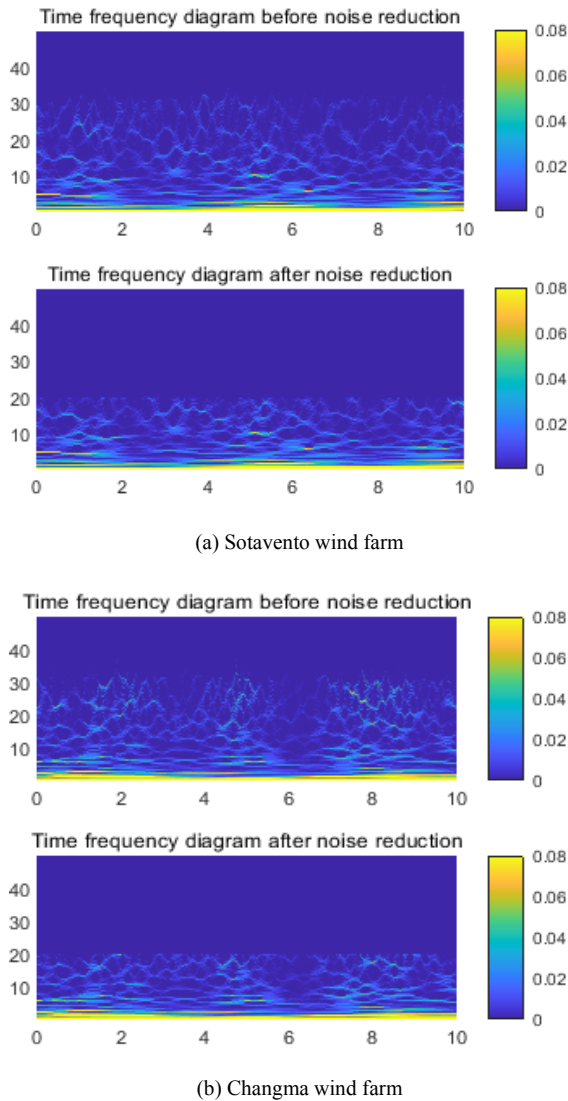


Figure 3. Time frequency diagram before and after SWT noise reduction

SWT is used to process the two sets of data, and the synchronous extrusion output of the signal is obtained and the bandpass filter is carried out. During processing, only a fixed frequency band is maintained, so as to eliminate the high frequency noise points, and SWT inverse transformation is performed on the output after noise reduction. The time-frequency diagram of signal SWT after noise reduction is shown in Figure 3. According to the figure above, noise above 25 Hz is eliminated. The data comparison after noise reduction is shown in Figure 4. The data is significantly smoother after noise reduction.

The use of SWT can reduce the noise signal interference, remove the impact of useless information, help to find and deeply mine data information, and improve the modeling efficiency.

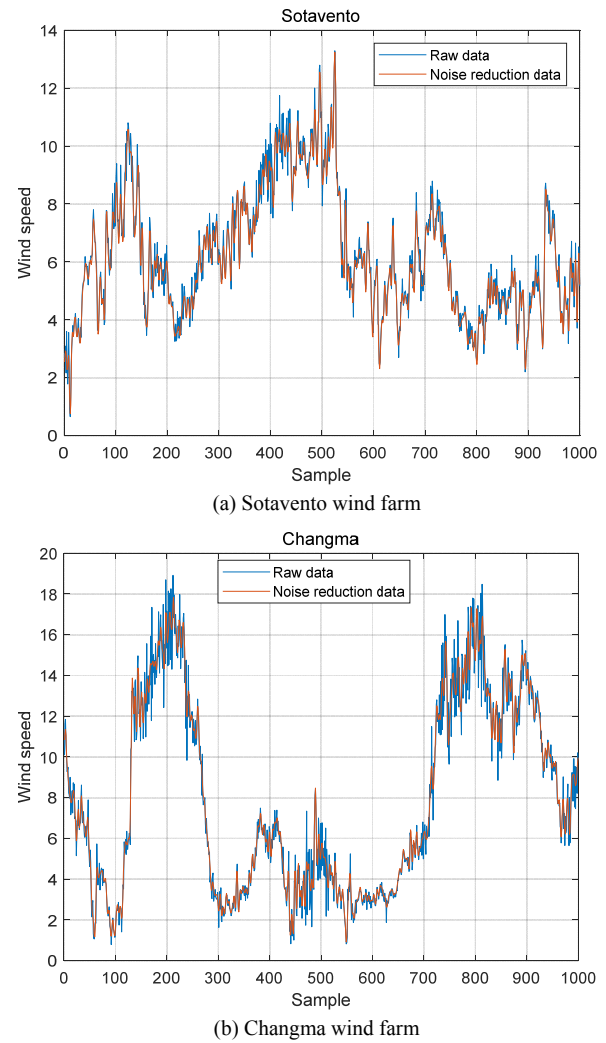


Figure 4. Data comparison before and after noise reduction

Reduce the waste of power resources and lay the foundation for follow-up work.

B. Establish Prediction Model

The key parameters of LSTM have a significant impact on the prediction performance of the model. When the model is built larger and the function is enhanced, the risk of overfitting increases.

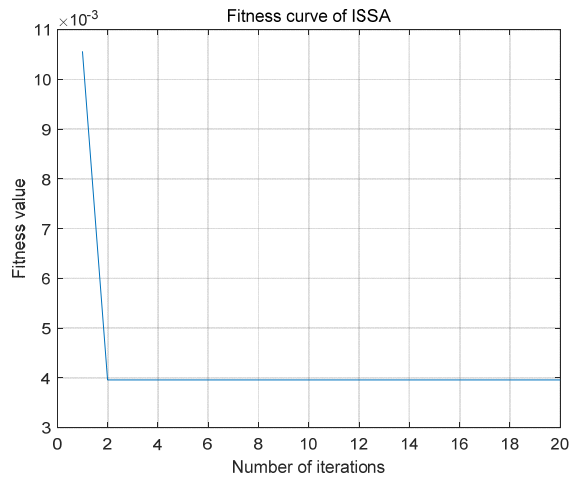
Manually adjusting the parameters of the LSTM not only increases the time cost, but also has a low probability of selecting parameters that can improve performance. It is prone to overfitting, which reduces the learning ability of the model.

For this reason, the optimization algorithm is used to optimize the parameters. While saving labor time and cost, it can effectively find the optimal parameter value that can enhance the performance of the prediction model, and further improve the stability of the hybrid model.

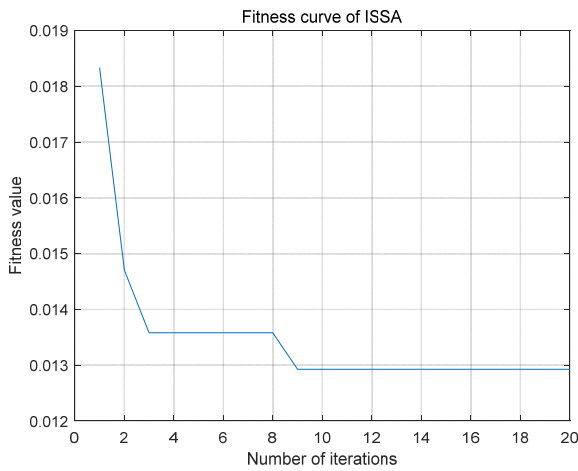
Based on the above noise reduction sequence for rolling modeling, the above noise-reduced sequences are used for rolling modeling. For LSTM's hyperparameter selection problem, the above enhanced Sparrow algorithm combining cauchy mutation and back learning was used for optimization. These hyperparameters include the number of hidden layer neurons, learning rate and iteration times. The following table shows the parameters involved in the algorithm.

TABLE II. SPECIFIC PARAMETER SETTINGS OF THE MODEL

Model	Parameters	Value
ISSA	population number	10
	proportion of discoverers	0.4
	proportion of scouts	0.2
	safety value	0.6
	number of iterations	20
LSTM	number of input nodes	10
	number of output nodes	1
	number of hidden layers	2
	Number of nodes in each hidden layer	20,20
	batchsize	16
	epochs	20
	learning rate	0.005



(a) Sotavento wind farm



(b) Changma wind farm

Figure 5. Fitness curve

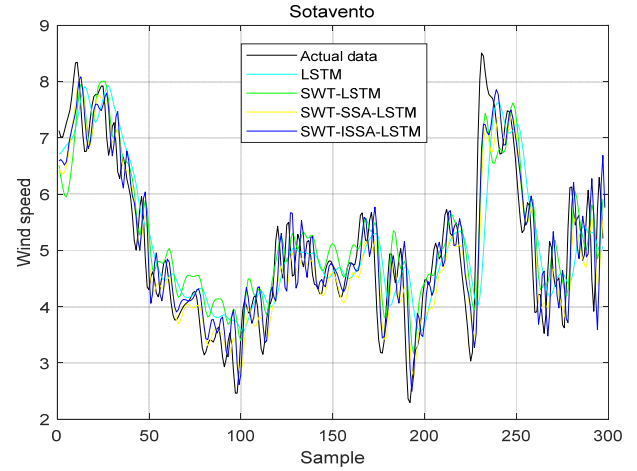
The Figure 5 shows the fitness curve of the optimization process. With the optimization process of sparrow search algorithm, a set of hyperparameters can be found, which makes the error of the network trained with this set of hyperparameters become lower. For this reason, the fitness curve is a declining curve.

The LSTM network was reconstructed using the optimized hyperparameters, and the LSTM model (SWT-ISSA-LSTM) based on SWT and ISSA optimization was obtained. The other three comparison models include: single

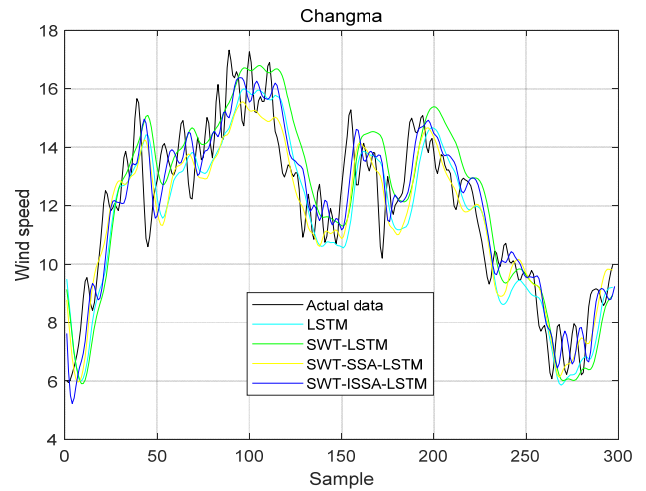
network model LSTM, a LSTM model based on SWT (SWT-LSTM) and an optimized LSTM model based on SWT and sparrow search algorithm (SWT-SSA-LSTM).

C. Evaluation of Prediction Results

The prediction effects of different models are shown in Figure 6, where the actual data is denoted by the black line, the cyan is the LSTM model, the green is the SWT-LSTM model, the yellow is SWT-SSA-LSTM, and the blue is SWT-ISSA-LSTM model.



(a) Sotavento wind farm



(b) Changma wind farm

Figure 6. Prediction results of different models on data sets

As can be seen from the figure, the blue line most closely conforms to the trend of the black line, indicating that the prediction effect of the proposed model is closest to the actual data. After SWT noise reduction, the redundant information can be effectively removed, the prediction accuracy can be improved by using SSA to optimize LSTM, and the improvement of integrating Cauchy variation and reverse learning also has obvious effect.

Table III lists the error analysis results of each model. It can be seen that the index values of the SWT-ISSA-LSTM model proposed in this paper are the smallest. Compared with LSTM model, SWT-LSTM model reduces the error, which shows that it is effective to remove interference by SWT. Using SWT-SSA-LSTM model for prediction, compared with SWT-LSTM and LSTM, the error continues to decrease, indicating that SSA can improve the prediction accuracy.

TABLE III. ERROR ANALYSIS OF PREDICTION RESULTS

Wind Farm	model	MAE	RMSE	MAPE(%)
Sotavento	LSTM	0.6391	0.8820	13.4064
	SWT-LSTM	0.5467	0.6929	12.0820
	SWT-SSA-LSTM	0.4656	0.6264	9.5616
	SWT-ISSA-LSTM	0.3039	0.3927	6.3519
Changma	LSTM	1.2998	1.6266	11.3958
	SWT-LSTM	1.1956	1.5092	10.7456
	SWT-SSA-LSTM	0.9594	1.2071	8.4061
	SWT-ISSA-LSTM	0.7597	0.9608	6.7425

Finally, compared with SWT-SSA-LSTM model, the MAPE values of SWT-ISSA-LSTM proposed in this paper in two wind farms are reduced by 3.2097% and 1.6591% respectively, which prove that the optimization combining Cauchy variation and reverse learning is meaningful. In each improvement process, the error value decreases, indicating that each optimization process is reasonable and effective.

V. CONCLUSION

The reliability of wind speed forecast is the guarantee of accurate and effective power grid scheduling, which is meaning a lot to the stable running of the system. In this paper, a prediction scheme using Cauchy variational and reverse learning is established. Firstly, SWT is used to denoise the volatile data to eliminate the interference of redundant information. Next, the LSTM network with long-term memory function is used for prediction, and optimized by SSA method. For the traditional SSA, the population diversity decreases in the late iteration, and it is easy to be locally optimal. In this paper, an ISSA algorithm combining Cauchy mutation and reverse learning is used to enhance the optimization ability. Experimental results demonstrate that the model can better improve the prediction quality and ensure that each step of optimization and improvement is reasonable and effective.

ACKNOWLEDGMENT

The authors thank the anonymous referees for the thoughtful and constructive suggestions that led to a considerable improvement of the paper.

REFERENCES

- [1] Y. Zhang, Y. Li, G. Zhang, "Short-term wind power forecasting approach based on Seq2Seq model using NWP data," *Energy*, 2020: 118371. doi:10.1016/j.energy.2020.118371
- [2] H. Shuai, X. Yue, H. Zhang, S. Xie, J. Li, C. Gu, W. Sun, J. Liu., "Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction," *Applied Energy*, 2021, 293: 116951. doi:10.1016/j.apenergy.2021.116951
- [3] E. Erdem, J. Shi. "ARMA based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, 2011, 88(4): 1405-1414. doi:10.1016/j.apenergy.2010.10.031
- [4] W. Li, X. Jia, X. Li., Y. Wang, J. Lee, "A Markov model for short term wind speed prediction by integrating the wind acceleration information," *Renewable Energy*, 2021, 164: 242-253. doi:10.1016/j.renene.2020.09.031
- [5] J. Zhou, J. Shi, G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Conversion and Management*, 2011, 52(4): 1990-1998. doi:10.1016/j.enconman.2010.11.007
- [6] D. Putz, M. Gumhalter, H. Auer, "A novel approach to multi-horizon wind power forecasting based on deep neural architecture," *Renewable Energy*, 2021, 178: 494-505. doi:10.1016/j.renene.2021.06.099
- [7] Y. Zhang, Y. Zhao, X. Shen, J. Zhang, "A comprehensive wind speed prediction system based on Monte Carlo and artificial intelligence algorithms," *Applied Energy*, 2022, 305: 117815. doi:10.1016/j.apenergy.2021.117815
- [8] S. Wang, N. Zhang, L. Wu, Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method" *Renewable Energy*, 2016, 94: 629-636. doi:10.1016/j.renene.2016.03.103
- [9] X. J.Chen, J. Zhao, X. Z. Jia, Z. L. Li, "Multi-step wind speed forecast based on sample clustering and an optimized hybrid system," *Renewable Energy*, 2021, 165: 595-611. doi:10.1016/j.renene.2020.11.038
- [10] S. Preitl, R. E. Precup, J. K. Tar, M. Takács, "Use of multi-parametric quadratic programming in fuzzy control systems," *Acta Polytechnica Hungarica*, 2006, 3(3): 29-43
- [11] R. E. Precup, R. C. David, R. C. Roman, A. I. Szedlak-Stinean, E. M. Petriu, "Optimal tuning of interval type-2 fuzzy controllers for nonlinear servo systems using Slime Mould Algorithm," *International Journal of Systems Science*, 2021: 1-16. doi:10.1080/00207721.2021.1927236
- [12] M. Moattari, M. H. Moradi, "Conflict monitoring optimization heuristic inspired by brain fear and conflict systems," *Int J Artif Intell*, 2020, 18(1): 45-62
- [13] C. Song, L. Yao, C. Hua, Q. Ni, "A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory," *Journal of Hydrology*, 2021, 603(Part A): 126879. doi:10.1016/j.jhydrol.2021.126879
- [14] Q. Mao, Q. Zhang, "Improved Sparrow algorithm integrating cauchy mutation and reverse learning," *Journal of Frontiers of Computer Science and Technology*, 2020, 15(6): 1155-1164. doi:10.3778/j.issn.1673-9418.2010032
- [15] G. Memarzadeh, F. Keynia, "A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets," *Energy Conversion and Management*, 2020, 213: 112824. doi:10.1016/j.enconman.2020.112824
- [16] Y. Zhang, G. Pan, Y. Zhao, Q. Li, F. Wang, "Short-term wind speed interval prediction based on artificial intelligence methods and error probability distribution," *Energy Conversion and Management*, 2020, 224: 1-14. doi:10.1016/j.enconman.2020.113346
- [17] A. Glowacz, "Ventilation diagnosis of angle grinder using thermal imaging" *Sensors* 2021; 21:2853. doi:10.3390/s21082853
- [18] Y. Zhao, W. Zhang, X. Gong, C. Wang, "A novel method for online real-time forecasting of crude oil price," *Applied Energy*, 2021; 303: 117588. doi:10.1016/j.apenergy.2021.117588
- [19] Y. Nie, N. Liang, J. Wang, "Ultra-short-term wind-speed bi-forecasting system via artificial intelligence and a double-forecasting scheme," *Applied Energy*, 2021, 301: 117452. doi:10.1016/j.apenergy.2021.117452
- [20] T. Liang, Q., Zhao, Q. Lv, H. Sun, "A novel wind speed prediction strategy based on Bi-LSTM, MOFADA and transfer learning for centralized control centers," *Energy*, 2021, 230: 120904. doi:10.1016/j.energy.2021.120904
- [21] B. Lin, C. Zhang, "A novel hybrid machine learning model for short-term wind speed prediction in inner Mongolia, China," *Renewable Energy*, 2021. doi:10.1016/j.renene.2021.07.126
- [22] W. Shuai, W. Jianzhou, H. Lu, W. Zhao, "A novel combined model for wind speed prediction-Combination of linear model, shallow neural networks, and deep learning approaches," *Energy*, 2021, 234: 121275. doi:10.1016/j.energy.2021.121275
- [23] Y. Zhang, G. Pan, "A hybrid prediction model for forecasting wind energy resources," *Environmental Science and Pollution Research*, 2020, 27(16): 19428-19446. doi:10.1007/s11356-020-08452-6
- [24] A. Glowacz, "Fault diagnosis of electric impact drills using thermal imaging," *Measurement* 2021; 171:108815. doi:10.1016/j.measurement.2020.108815
- [25] T. B. M. J. Ouarda, C. Charron, "Non-stationary statistical modelling of wind speed: A case study in eastern Canada," *Energy Conversion and Management*, 2021, 236: 114028. doi:10.1016/j.enconman.2021.114028