



ARTICLE

Short-Term Wind Power Prediction Based on WVMD and Spatio-Temporal Dual-Stream Network

Yingnan Zhao*, Yuyuan Ruan and Zhen Peng

School of Computer and Science, Nanjing University of Information Science and Technology, Nanjing, 210044, China

*Corresponding Author: Yingnan Zhao. Email: zh_yingnan@126.com

Received: 17 July 2024 Accepted: 23 August 2024 Published: 15 October 2024

ABSTRACT

As the penetration ratio of wind power in active distribution networks continues to increase, the system exhibits some characteristics such as randomness and volatility. Fast and accurate short-term wind power prediction is essential for algorithms like scheduling and optimization control. Based on the spatio-temporal features of Numerical Weather Prediction (NWP) data, it proposes the WVMD_DSN (Whale Optimization Algorithm, Variational Mode Decomposition, Dual Stream Network) model. The model first applies Pearson correlation coefficient (PCC) to choose some NWP features with strong correlation to wind power to form the feature set. Then, it decomposes the feature set using Variational Mode Decomposition (VMD) to eliminate the non-stationarity and obtains Intrinsic Mode Functions (IMFs). Here Whale Optimization Algorithm (WOA) is applied to optimise the key parameters of VMD, namely the number of mode components K and penalty factor a . Finally, incorporating attention mechanism (AM), Squeeze-Excitation Network (SENet), and Bidirectional Gated Recurrent Unit (BiGRU), it constructs the dual-stream network (DSN) for short-term wind power prediction. Comparative experiments demonstrate that the WVMD_DSN model outperforms existing baseline algorithms and exhibits good generalization performance. The relevant code is available at <https://github.com/ruanyuyuan/Wind-power-forecast.git> (accessed on 20 August 2024).

KEYWORDS

Wind power prediction; dual-stream network; variational mode decomposition (VMD); whale optimization algorithm (WOA)

1 Introduction

Due to its low cost, abundance, and environmental friendliness, wind energy has gained global attention [1,2]. However, the inherent volatility, intermittence, and uncertainty of wind power pose significant challenges to the safe and stable operation of power systems. Therefore, accurate wind power prediction is crucial for improving the reliability and safety of system operations [3].

Currently, wind power prediction methods are mainly divided into four categories: physical methods, statistical methods, traditional machine learning methods and deep learning methods. Although the first three traditional ones have certain applications in wind power prediction, there are also some limitations and challenges. The physical method has high requirements on the accuracy



of meteorological factors, and it is difficult to cope with the complex and changeable wind field conditions [4]. The statistical method lacks the ability to model nonlinear relationships and has poor adaptability to complex wind farm dynamic changes [5]. Traditional machine learning performs poorly on nonlinear and non-stationary data and has limited model generalization ability [6]. In contrast, deep learning technology has great advantages and has become the current mainstream method [7].

In wind power prediction using deep learning, initial approaches employed single models such as LSTM [8], CNN [9,10], GRU [11], and BiGRU [12]. However, wind power data consist of sequential information that includes both temporal and spatial components [13]. While single models effectively handle temporal data from time series of wind power, they often overlook spatial features. Spatial information in this context encompasses different geographical locations [14] and diverse meteorological factors [15]. To address these spatio-temporal characteristics, hybrid models have been developed for wind power prediction using deep learning techniques. These include CNN-LSTM [16], CNN-BiLSTM [17], CNN-BiLSTM-AM [18], etc. Although the above hybrid models can effectively deal with spatio-temporal features, they also increase the complexity. The two-stream network applied to action recognition [19] presented recently, uses parallel processing to handle spatio-temporal features and shows superior performance than the hybrid models. Therefore, it is applied to some fields rapidly. For example, Reference [20] proposes a two-stream three-dimensional CNN, which captures spatial and temporal information on a fine time scale for skeletal-based action recognition. Reference [21] proposes a DSN composed of CNN and GRU to learn spatio-temporal features from photovoltaic power generation and meteorological data, which improved the accuracy of prediction. Thus, it is natural to adopt it in the field of wind power prediction. However, how to improve the construction of DSN is still a challenge. Generally, one branch of the DSN processes spatial data while the other handles temporal data. What types of networks are used for each branch, and how many are there? The features generated by these branches may have some redundancy. How can this redundancy be reduced? These are some of the challenges faced in this field.

Additionally, to address the intermittency, volatility, and uncertainty of wind power generation, the aforementioned models often incorporate data preprocessing methods such as empirical mode decomposition (EMD) [22], ensemble empirical mode decomposition (EEMD) [23], wavelet transform (WT) [24], and variational mode decomposition (VMD) [25]. Among them, the VMD technology shows the best performance and noise robustness [16,25]. However, one of the drawbacks of this method is that the number of mode components K and penalty factor a have a large impact on the decomposition results [26]. Hence, it is important to consider the optimal combination of key parameters K and a of VMD.

Motivated by the basic ideas mentioned above, the paper proposes a superior wind power named WVMD-DSN. It first utilizes the PCC to choose some NWP features to form the feature set, Then, it decomposes the feature set using VMD, while adopting WOA to fine-tune the number of mode components K and penalty factor a . Finally, it constructs the DSN for short-term wind power prediction, integrating AM, SENet, and BiGRU.

The main contributions of this paper are as follows:

- 1) It proposes a novel dual-stream network model and applies it to wind power generation, achieving better prediction results than other baseline models and demonstrating strong generalization performance.
- 2) In order to enhance the prediction accuracy and address potential errors, it introduces WOA to design the key parameters of VMD adaptively.

2 Background Technologies

In this chapter, we first discuss the theoretical foundation of VMD preprocessing technology and the evaluation indices of envelope entropy and mutual information used for WOA-optimized VMD. We then explain the theoretical underpinnings of SENet and BiGRU as employed in the DSN network model.

2.1 VMD and WOA

VMD is an adaptive signal decomposition method that employs a non-recursive decomposition scheme [27]. It decomposes signals into a specified number of IMFs with different center frequencies. Applying the VMD decomposition preprocessing method to wind power involves the following constrained variational model:

$$\begin{aligned} \min_{|u_k|, |\omega_k|} & \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} & \sum_{k=1}^K u_k(t) = \tilde{X}(t) \end{aligned} \quad (1)$$

In the equation, $\tilde{X}(t)$ is the wind power generation series; K is the number of modes to be decomposed; $\delta(t)$ is the Dirac function; j is the imaginary unit; t is time; $u_k(t)$ is the IMF component at time k ; ω_k is the central frequency corresponding to the IMF component; $*$ is convolution operation.

To solve Eq. (1), the Lagrange multiplication operator λ is introduced to transform the constrained problem into a non-constrained problem and obtain the augmented Lagrange expression.

$$\begin{aligned} L(u_k, \omega_k, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \quad (2)$$

In Eq. (2), α is the penalty factor, which is used to reduce the influence of Gaussian noise. The parameters u_k , ω_k and λ are iteratively updated by the alternate direction method of the multiplier.

WOA [28] is a swarm intelligence optimization algorithm that effectively prevents getting trapped in local minima, unlike Particle Swarm Optimization (PSO) [29] and Grey Wolf Optimizer (GWO) [30], thereby enhancing global optimization efficiency. Therefore, it is suitable to solve some complex, changeable and nonlinear problems. Mutual Information (MI) primarily used in information theory to quantify the degree of correlation between two events, is less susceptible to external interference [31]. Here, we adopt WOA to optimize the key parameters K and α of VMD, and choose the envelope entropy and MI as the fitness function. The size of the envelope entropy reflects the uncertainty of the probability distribution. The larger the entropy, the greater the uncertainty of the signal. The calculation of the envelope entropy E_p of signal $\tilde{X}(t)$ is as follows:

$$\begin{cases} E_p = - \sum_{t=1}^N p_t \lg p_t \\ p_t = \frac{a(t)}{\sum_{t=1}^N a(t)} \end{cases} \quad (3)$$

Whereas, N is the number of sampled points of the signal, $a(t)$ is the envelope signal obtained through Hilbert demodulation of signal $\tilde{X}(t)$, and p_i represents the normalized result of signal $a(t)$.

MI expression is as follows:

$$MI(X, Y) = H(Y) - H(Y|X) \quad (4)$$

In the equation, X and Y represent different events, $H(X)$ is the entropy of Y , $H(Y|X)$ is the conditional entropy of Y given X . The normalized mutual information is processed as follows:

$$mi = MI(X, Y) / \sqrt{H(X) \cdot H(Y)} \quad (5)$$

A higher mutual information entropy value indicates a stronger correlation between two events. For IMF components, those that better capture the characteristic information of the wind power signal will have a larger mutual information entropy value. The composite index established in this article is:

$$I = E_p * mi \quad (6)$$

The composite index considers both the noise level and characteristic information of IMF components. When an IMF component contains less noise and richer characteristic information, the composite index I has a smaller value. Therefore, the minimum value of I is adopted as the fitness function, expressed as:

$$fitness = \min_{1 \sim K} \{I\} \quad (7)$$

2.2 SENet and BiGRU

SENet [32] is a CNN-based attention mechanism for learning feature relationships between different channels. By introducing the squeeze-excitation (SE) module, it learns the relationship between the CNN convolution kernel's channels, and employs the channel AM to adaptively recalibrate channel-wise feature responses by explicitly modeling the interdependencies. SEblock adjusts the importance of each convolutional channel by learning their weights, thereby enhancing the network's perception of different feature channels. The formula is as follows:

$$z_c = \frac{1}{h \times w} \sum_{\substack{1 \leq i \leq h \\ 1 \leq j \leq w}} u_c(i, j) \quad (8)$$

$$s = \sigma(w_2 \delta(w_1 z)) \quad (9)$$

$$\tilde{x}_c = s_c \cdot u_c \quad (10)$$

In Eq. (8), z_c is the squeeze output, and u_c is the eigenvalue of the characteristic graph in channel c . In Eq. (9), w_2 and w_1 are fully connected layer operation, and δ denotes a ReLU layer operation. And then the output go through the sigmoid function σ to get s which is to characterize the weights of feature maps in matrix u_c . In Eq. (10), s_c reflects the importance weight of each feature channel, and the weight coefficient of each channel can be learned through $s_c \cdot u_c$.

As for BiGRU, it is on the base of the traditional GRU [33] network, which the transmission of information is unidirectional. The mathematical model is shown below:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (11)$$

$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t] + b_z) \quad (12)$$

$$\tilde{h}_t = \tanh (W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (13)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (14)$$

In the equations, W_r , W_z and W_h are the weights of the internal parameters of the GRU. b_r , b_z and b_h are the biases of the internal parameters of the GRU. σ is the sigmoid function, and \tanh is the hyperbolic tangent function.

In the wind power forecasting, the current output relates not only to the state at the subsequent time but also to the previous. The input information at each time step is fed into two GRU networks in opposite directions in a BiGRU. Therefore, the BiGRU model can capture the strong temporal correlations in wind power sequences effectively.

3 WVMD_DSN Model

The proposed WVMD_DSN is illustrated in Fig. 1, which primarily consists of the following steps:

- (1) Clean the data, and then split the data into NWP and wind power data.
- (2) Decompose the wind power by VMD and get the IMF components. Here VMD is optimized by WOA to select the best mode number K and penalty factor.
- (3) Using PCC to calculate the correlation between NWP data and power features, and choose those with stronger correlations to form the feature set NWP_{pcc} .
- (4) Integrating the NWP_{pcc} and IMF components to form the input sets.
- (5) Predict each value by DSN, which integrates SENet, BiGRU, and AM.
- (6) Combine all the predicted values to form the final value.

3.1 Data Preprocessing

The study utilizes spatio-temporal data consisting of Wind Power, Wind Speed-10 (WS-10), Wind Speed-30 (WS-30), Wind Speed-50 (WS-50), Wind Direction-10 (WD-10), Wind Direction-30 (WD-30), Wind Direction-50 (WD-50), Humidity, Temperature and Atmospheric Pressure. Here, Wind Speed-10 represents the wind speed at 10 m on the wind measurement tower, and the other numeric features follow the same meanings. The data preprocessing methods include the mean imputation for missing values and the MinMax scaling technique to normalize each attribute to the range of 0 to 1. Further, it employs PCC to select highly correlated NWP features. Additionally, to reduce the non-stationary characteristics of wind power, VMD is employed for decomposition, obtaining the IMF components. This is illustrated in Fig. 2, where K represents the number of IMFs.

After preprocessing the data, we employ a sliding window approach to select temporal data and prediction labels for the model. We use data from the first two hours as input, with a sample resolution of fifteen minutes. The sliding window size is set to 8, and the label window corresponds to the next fifteen minutes, with a prediction step of 1. Finally, the dataset is divided into training, validation, and test sets.

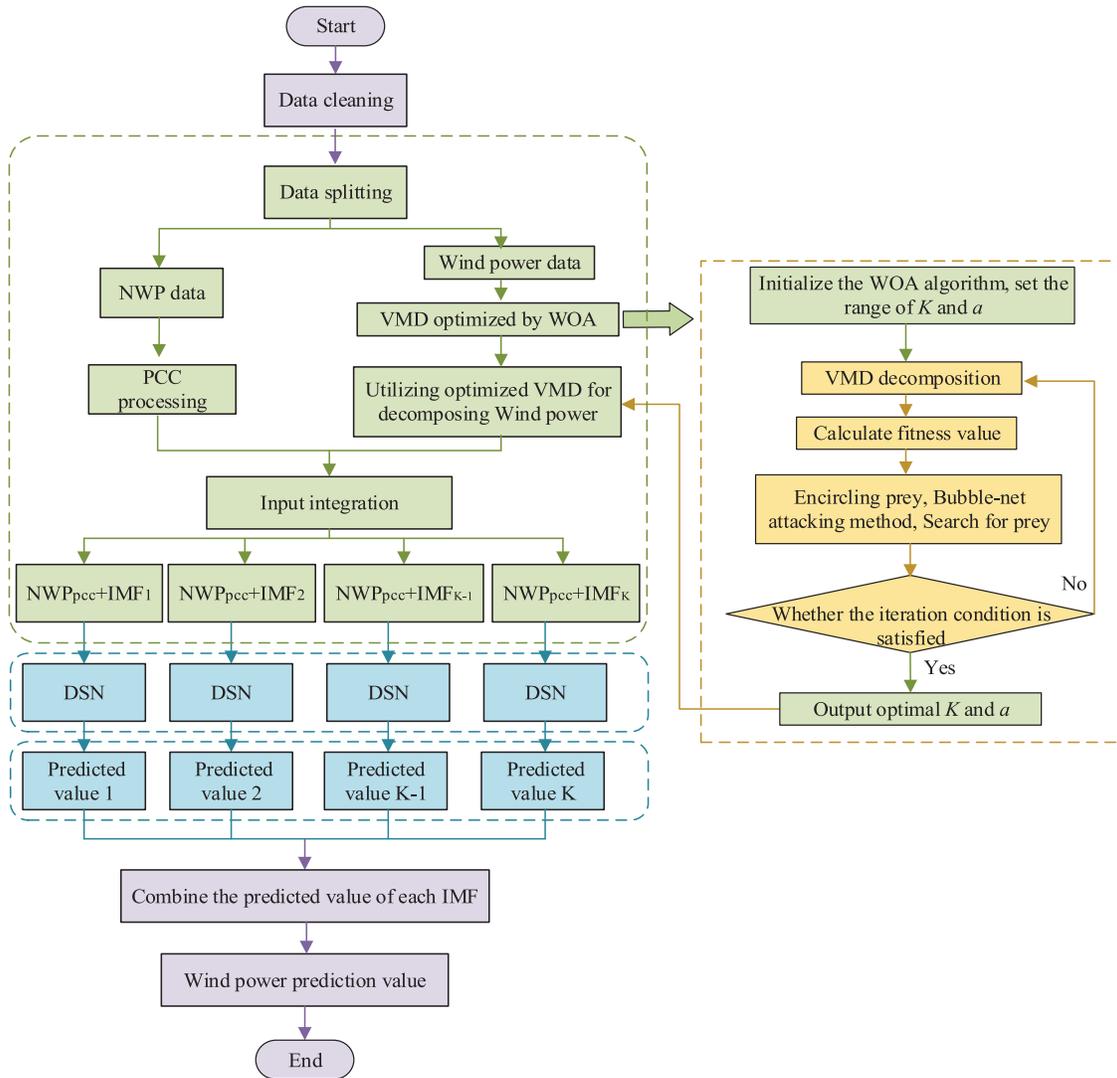


Figure 1: The architecture of WVMD_DSN

3.2 DSN Structure

In the paper, the proposed DSN consists of SENet and BiGRU, using for parallel processing of spatio-temporal features, as shown in Fig. 3. It extracts irregular trends and complex features from the variables of historical wind power data. The first stream uses SENet, which alone cannot capture spatio-temporal information. Therefore, we introduce a second stream using BiGRU. The SENet stream processes spatial data from NWP, including variables such as temperature, humidity, wind speed, and IMF components from VMD. The BiGRU stream focuses on extracting time series data of wind power also processed by VMD. The BiGRU stream mainly extracts the time series data of wind power processed by VMD.

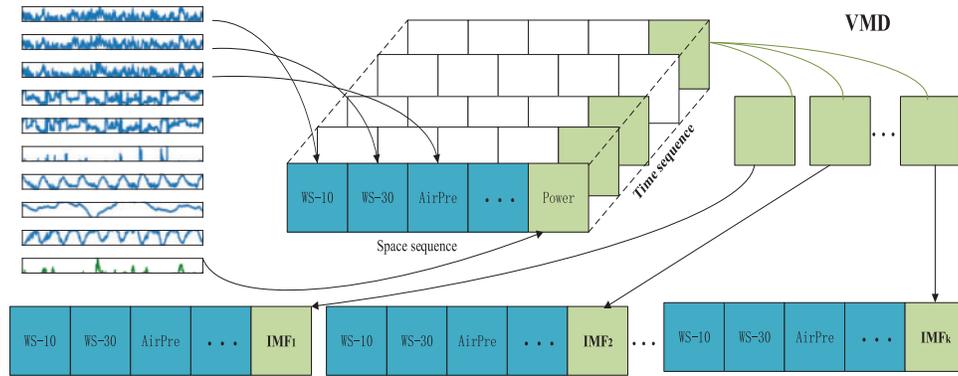


Figure 2: Schematic diagram of spatio-temporal wind power data processing based on VMD decomposition

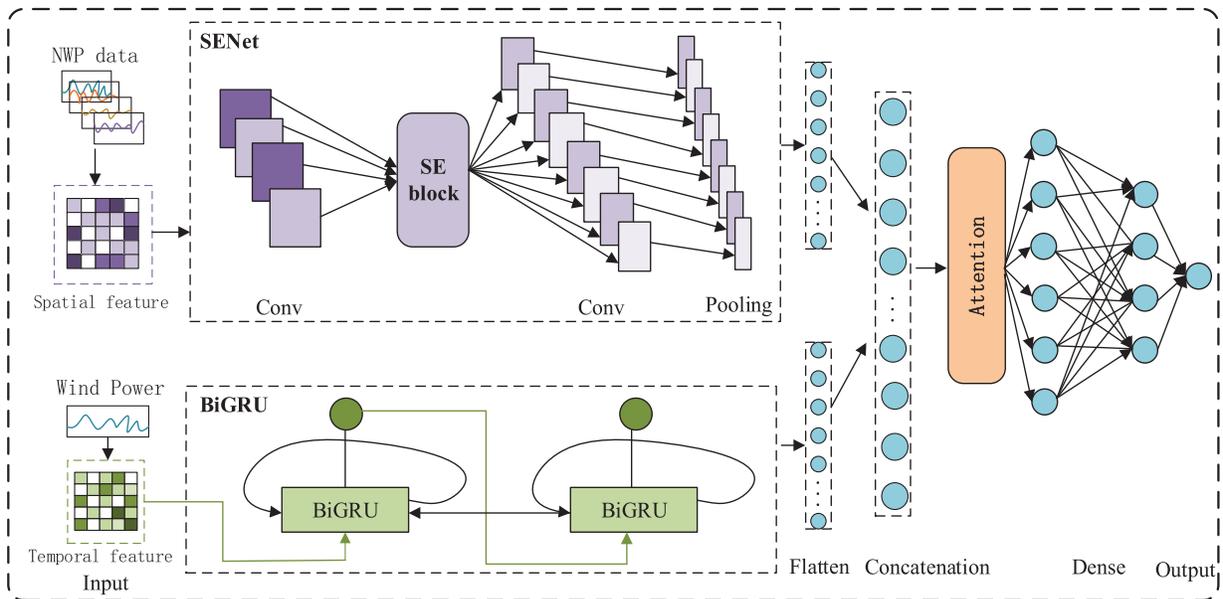


Figure 3: DSN network structure

The DSN includes an input layer, multiple hidden layers, and an output layer. The SENet stream’s hidden layers consist of convolutional layers, SEblocks, and pooling layers, which extract spatial features from various meteorological factors influencing wind power generation, that is the NWP data. The BiGRU stream comprises two BiGRU layers, which capture the temporal features of wind power generation. Then, the outputs of both streams are concatenated to form a single feature vector, which is passed to an attention module to extract more valuable output results. It produces the ultimately wind power prediction result.

4 Experiments and Results Analysis

In this chapter, we first describe the attributes of the two wind power datasets used in this paper. Next, we detail the WVMd preprocessing technique and the parameters employed in the DSN model.

We then assess the efficacy of WVMD denoising by comparing the processed signals with Gaussian white noise and a normal distribution. Following this, WVMD is applied to wind power data. Finally, we perform ablation experiments to analyze the parallel computing time of the DSN model, compare WVMD_DSN with the current baseline model, and conduct a generalization analysis using wind power dataset 2. This comprehensive approach demonstrates the effectiveness of the WVMD_DSN method in wind power prediction.

4.1 Datasets

To evaluate the proposed WVMD_DSN, it uses the actual power and weather forecast numerical data collected from a region in Xinjiang, China. The data sampling period spans from 1 January, 2019 to December 31, 2019, with a sampling frequency of 15 min. Thus it consists in a total of 35,040 data points. Wind power generation dataset includes historical wind power and NWP data, that is, timestamp, Wind Power (WP), Temperature (Tem), Air Pressure (AP), Humidity (Hum), Wind Direction (WD), and Wind Speed (WS). To further investigate the generalization ability of the WVMD_DSN model, wind power dataset 2 is adopt. This dataset originated from Inner Mongolia, China. The specific features of wind power generation include timestamps, Air Density (AD), Temperature (Tem), Air Pressure (AP), Humidity (Hum), Wind Speed (WS), Wind Direction (WD), and Wind Power (WP). The sample information of the two datasets are shown in [Table 1](#).

Table 1: Statistical information of data attributes

| Datasets | Attributes | Mean | Std | Min | 25% | 50% | 75% | Max |
|-------------------------|----------------------------|---------|--------|--------|--------|---------|---------|---------|
| Wind power dataset 1 | Wind power (MW) | 53.90 | 56.73 | 0.258 | 3.526 | 29.037 | 99.741 | 202.22 |
| | Temperature (°C) | 11.324 | 14.03 | -18.66 | -1.576 | 13.105 | 23.04 | 40.13 |
| | Air pressure (hPa) | 888.509 | 5.793 | -874.5 | 883.63 | 888.638 | 892.708 | 905.307 |
| | Humidity (%) | 33.59 | 20.76 | 2.506 | 16.647 | 28.122 | 47.544 | 94.92 |
| | Wind direction-10 m (°) | 156.22 | 86.821 | 0.00 | 82.25 | 137.966 | 230.509 | 359.946 |
| | Wind direction-30 m (°) | 158.638 | 89.835 | 0.00 | 79.707 | 137.621 | 239.895 | 360.0 |
| | Wind direction-50 m (°) | 11.587 | 16.812 | 0.00 | 5.852 | 6.392 | 11.539 | 355.823 |
| | Wind speed-10 m (m/s) | 4.818 | 3.516 | 0.00 | 2.364 | 3.869 | 6.602 | 19.246 |
| | Wind speed-30 m (m/s) | 5.347 | 3.836 | 0.00 | 2.62 | 4.481 | 7.312 | 21.056 |
| Wind speed-50 m (m/s) | 5.774 | 4.054 | 0.00 | 2.772 | 4.66 | 8.000 | 22.382 | |
| Wind power dataset 2 | Air density (°C) | 1.106 | 0.053 | 0.981 | 1.062 | 1.105 | 1.147 | 1.246 |
| | Temperature (°C) | 1.537 | 12.309 | -28 | -8 | 1 | 12 | 32 |
| | Air pressure (hPa) | 86.95 | 0.55 | 85.38 | 86.6 | 86.9 | 87.4 | 88.7 |
| | Humidity (%) | 56.51 | 22.36 | 5.00 | 39.2 | 57.3 | 75.4 | 100.5 |
| | Wind speed-weather (m/s) | 7.377 | 3.197 | 0.00 | 5.00 | 7.20 | 9.6 | 20.8 |
| | Wind direction-weather (°) | 243.9 | 81.11 | 0.00 | 214.0 | 270.0 | 297.0 | 359.0 |
| | Wind speed-farm (m/s) | 8.057 | 3.847 | 0.00 | 5.190 | 7.650 | 10.680 | 26.77 |
| Wind direction-farm (°) | 202.04 | 105.94 | 3.00 | 89.00 | 232.00 | 294.00 | 359.00 | |
| | Wind Power (MW) | 20.205 | 15.610 | 0.00 | 5.46 | 17.850 | 34.450 | 50.230 |

4.2 Experimental Parameter Settings

The experiments in this study are conducted using Python 3.7 and the TensorFlow framework. They are performed on a server equipped with an AMD Ryzen 9 5900HX CPU @ 4.6 GHz, an NVIDIA GeForce RTX 4090 GPU, and 24 GB of RAM.

The WVMD_DSN model uses the Adam optimization algorithm with training epochs set to 100. [Table 2](#) illustrates the parameter settings for the WOA optimization algorithm and the parameter settings for the DSN dual-stream network structure.

Table 2: WVMD_DSN model parameter settings

| Type | Parameter |
|---------------------|-------------------------------------------------------------------------------|
| Convolutional layer | Filter: 8; Kernel size: 1 |
| SE block layer | Ratio: 2 |
| Convolutional layer | Filter: 16; Kernel size: 3 |
| Max pooling layer | Pooling size: 2 |
| BiGRU layer | Hidden units: 32 |
| BiGRU layer | Hidden units: 16 |
| Attention layer | Units: 96 |
| WOA | Population size: 20; Maximum number of iterations: 40 |
| VMD | Initial (K, a) value is $(7, 500)$; Optimized (K, a) value is $(9, 619)$ |

4.3 WVMD Processing

4.3.1 Analysis of WVMD Decomposition Techniques

In this section, we compare the effectiveness of the WVMD technique with VMD and WOA-VMD. The comparison includes simulated signals with Gaussian white noise and signals with normally distributed noise. Simulated Signal 1 contains four distinct frequency components: a_1 , a_2 , a_3 , and a_4 , sampled at 500 Hz. The original signal is a superposition of a_1 , a_2 , a_3 , and a_4 . Signal 1 with noise includes Gaussian white noise with a signal-to-noise ratio of 10 dB, while Signal 2 with noise contains normally distributed noise with a signal-to-noise ratio of 15 dB. The [Formula \(15\)](#) as follows:

$$\begin{cases} t_i = \frac{30\pi}{500} \cdot i, i \in [0, 499] \\ a_1 = 1.2 \cos(0.1\pi t_i + \pi/4) \\ a_2 = 0.8 \sin(0.3\pi t_i + \pi/6) \\ a_3 = 1.5 \cos(0.5\pi t_i + \pi/3) \\ a_4 = 1.0 \sin(0.7\pi t_i + \pi/2) \\ s = a_1 + a_2 + a_3 + a_4 \\ \xi = \text{noise } 1 \\ y = a_1 + a_2 + a_3 + a_4 + \xi \end{cases} \quad \begin{cases} t_i = \frac{30\pi}{500} \cdot i, i \in [0, 499] \\ a_1 = 1.5 \cos(0.2\pi t_i + \pi/3) \\ a_2 = 1.0 \sin(0.5\pi t_i + \pi/4) \\ a_3 = 2.0 \cos(0.8\pi t_i + \pi/6) \\ a_4 = 1.2 \sin(1.2\pi t_i + \pi/2) \\ s = a_1 + a_2 + a_3 + a_4 \\ \xi = \text{noise } 2 \\ y = a_1 + a_2 + a_3 + a_4 + \xi \end{cases} \quad (15)$$

To verify the WVMD decomposition and denoising capabilities, we denoise the simulated signal by WVMD, WOA-VMD and VMD separately. The parameter combination $[K, \alpha]$ for VMD is $[7, 500]$. In Signal 1, the values of $[K, \alpha]$ for WOA-VMD are $[3, 140]$, with a fitness value of 8.5471, and for

WVMD are [10, 916], with a fitness value of 0.8965. In Signal 2, the values of $[K, \alpha]$ for WOA-VMD are [9, 532], with a fitness value of 8.6098, and for WVMD are [9, 334], with a fitness value of 1.1014. As shown in Table 3, the WVMD decomposition method outperforms WOA-VMD and VMD. Specifically, the WVMD-denoised signal has the highest SNR and R values and the lowest RMSE value. In both signals, compare the methods VMD and WOA-VMD, WVMD shows superior performance.

Table 3: Evaluation metrics for denoising simulated signals

| Simulated signal | Decomposition method | RMSE | R | SNR |
|-------------------------------------|----------------------|---------------|---------------|----------------|
| Gaussian white noise Signal 1 | VMD | 0.4208 | 0.9660 | 11.7269 |
| | WOA-VMD | 0.3682 | 0.9742 | 12.8857 |
| | WVMD | 0.3336 | 0.9787 | 13.7432 |
| Normally distributed noise Signal 2 | VMD | 0.3810 | 0.9840 | 14.7885 |
| | WOA-VMD | 0.3143 | 0.9886 | 16.4613 |
| | WVMD | 0.2973 | 0.9899 | 16.9435 |

4.3.2 Decomposition Results Based on WVMD

To further improve the data quality and eliminate the effects of noise, WVMD is adopted to decompose the wind power dataset 1. The number of whales is set to 20, the maximum number of iterations is 40, the number of variables is 2, the penalty factor is [100, 2000], and the K value range is [2, 10] and includes only integers. From Fig. 4a, it can be seen that the whale algorithm is gradually stable after twelve iterations, and the optimal fitness function is 2.140759. The optimization curve of the penalty factor a is shown in Fig. 4c. After twelve iterations, the optimal penalty parameter is 619. From Fig. 4b, the optimal number for the IMF is 9. Fig. 5b shows the spectra of each IMF component. The time domain of modal components obtained by WVMD decomposition is presented in Fig. 5a.

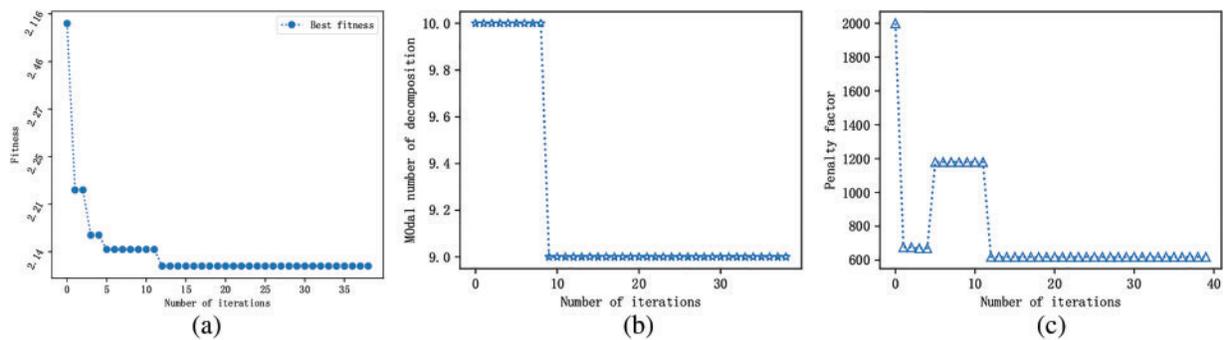


Figure 4: (a) Fitness curve of WVMD application in wind power; (b) Mode components K ; (c) Penalty factor a

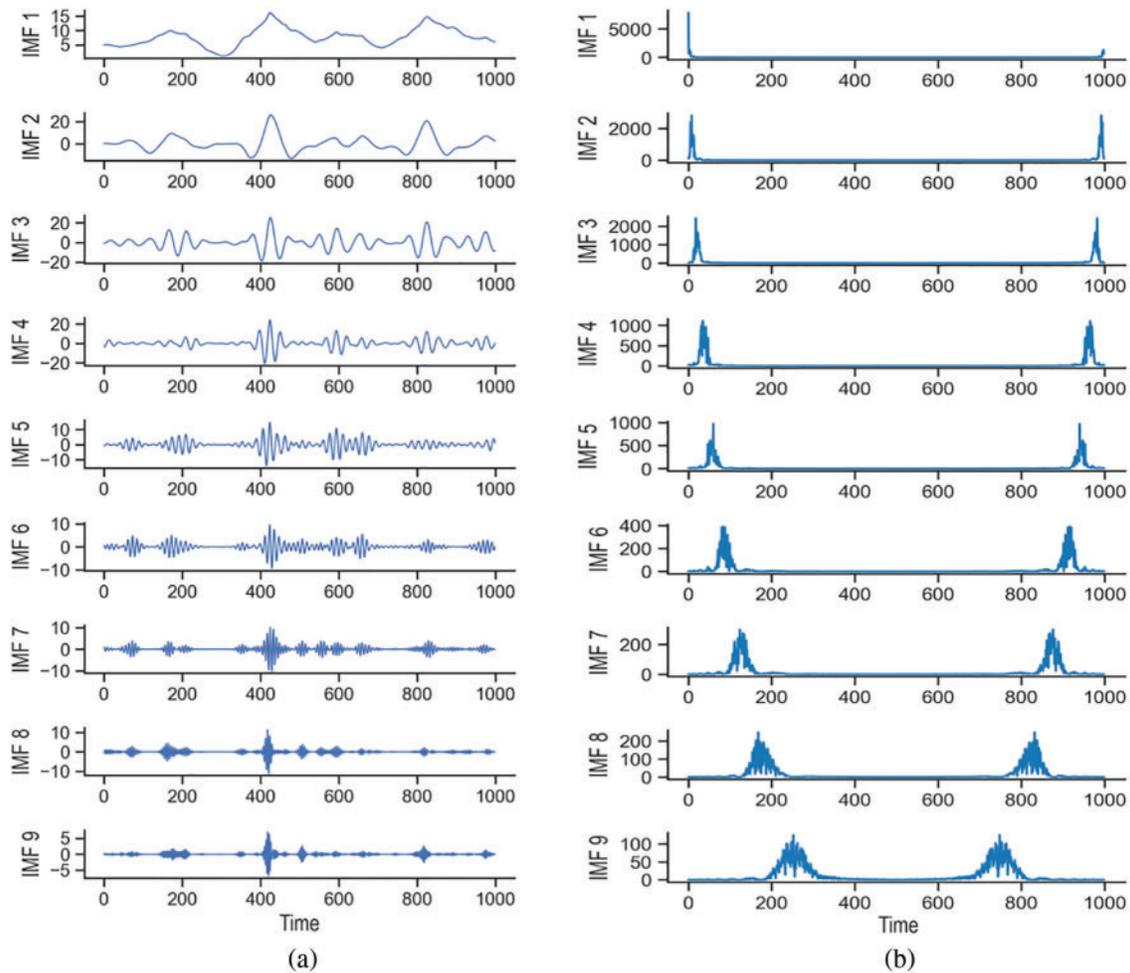


Figure 5: (a) WVMD decomposition of wind power; (b) Spectra of each IMF component

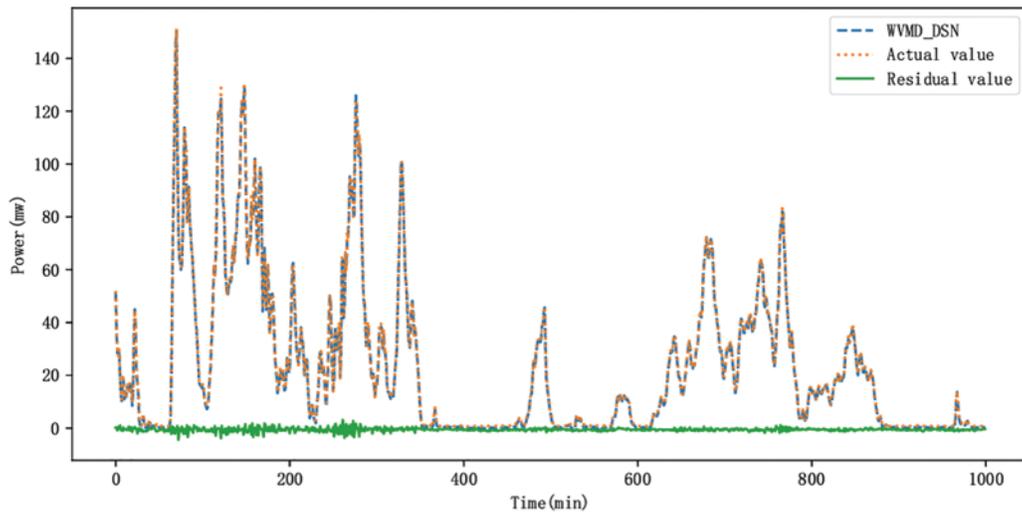
4.4 Comparison Analysis of WVMD_DSN

4.4.1 Ablation Study

In this section, using wind power generation datasets as an example, we validate the rationality of the WVMD_DSN model construction. As shown in Table 4, compared to the single models BiGRU and SENet, the hybrid model SENet-BiGRU shows an increase in RMSE and MAE by 11% and 11.5%, respectively. Similarly, compared to the hybrid model, the dual-stream model SENet-BiGRU exhibits an increase in RMSE and MAE by 4% and 8.3%, demonstrating the superior effectiveness of the DSN model over single and hybrid models. Incorporating WVMD as a preprocessing method before the DSN model notably enhances prediction accuracy, with RMSE, MAE, and NRMSE reaching as low as 1.1453, 0.8985, and 0.0056, respectively. Finally, the WVMD_DSN model, enhanced with attention mechanisms, achieves optimal performance with an RMSE of 1.0199 and an R^2 value of 99.95%. The predictive performance of the WVMD_DSN method is illustrated in Fig. 6, showing that the predicted residuals consistently fluctuate around the zero scale line of the original wind power sequence.

Table 4: Ablation study results

| Model | RMSE | MAE | NRMSE | R ² |
|----------------------------|---------------|---------------|---------------|----------------|
| BiGRU | 7.9515 | 4.7374 | 0.0394 | 0.9693 |
| SENet | 8.1671 | 4.4482 | 0.0404 | 0.9676 |
| SENet-BiGRU (Hybrid model) | 7.0702 | 4.1940 | 0.0350 | 0.9757 |
| SENet-BiGRU (Dual-stream) | 6.7830 | 3.8480 | 0.0336 | 0.9776 |
| VMD-Dual-stream | 4.2683 | 2.2497 | 0.0211 | 0.9911 |
| WOA-VMD-Dual-stream | 1.1453 | 0.8985 | 0.0056 | 0.9993 |
| WVMD_DSN | 1.0199 | 0.7298 | 0.0050 | 0.9995 |

**Figure 6:** Prediction results and residual plot of WVMD_DSN

4.4.2 Comparison with Baseline Model

Comprehensive Analysis of the DSN Model

In this section, the proposed DSN model is analyzed in detail. According to [Table 5](#), CNN has a training time of 94.5, making it the fastest among the models compared. CNN typically outperforms GRU and LSTM in processing sequence data because it uses parallel computing for feature mapping, rather than sequentially processing each element like recurrent neural networks. However, the RMSE and MAE of CNN alone are 9.2461 and 5.4185, respectively, indicating relatively poor prediction performance.

Table 5: Analysis of time efficiency of the DSN model

| Model | Training time/s |
|----------|-----------------|
| GRU [11] | 213.7 |
| LSTM [8] | 205.5 |

(Continued)

Table 5 (continued)

| Model | Training time/s |
|-----------------|-----------------|
| CNN [9] | 94.5 |
| BiGRU [14] | 266.5 |
| CNN-BiLSTM [18] | 194.9 |
| CNN-BiGRU [17] | 186.2 |
| SENet-BiGRU | 214.5 |
| DSN | 170.0 |

To address this, CNN is combined with BiGRU, CNN-BiLSTM, and other models. The CNN-BiGRU model has a shorter training period and higher prediction accuracy than BiGRU. This improvement is due to CNN's ability to effectively extract features from input data, especially when dealing with spatially structured data. By incorporating CNN, the output feature maps become more concise and representative, which reduces the data and complexity that BiGRU needs to handle. Although combining CNN with SENet and BiGRU improves prediction performance, it also increases time complexity. This is mainly because SENet includes an attention mechanism, which adds more parameters and complexity to the model.

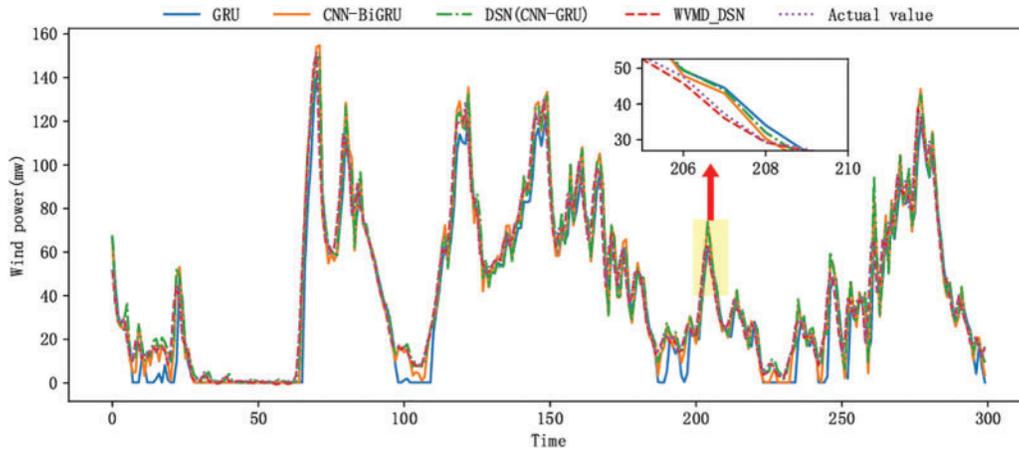
Compared to GRU, BiGRU, and CNN-BiGRU, the DSN model shows improved prediction accuracy. However, its training cycle increases by 20.4%, 36.1%, and 8.6%, respectively. RMSE rises by 21.1%, 14.6%, and 8.3%, while MAE increases by 31.6%, 18.7%, and 16.6%. Overall, DSN offers advantages in terms of prediction accuracy despite a longer training time.

Compare the Proposed WVMD_DSN Model

We compare the proposed WVMD_DSN model with other commonly used similar algorithms. The specific results are detailed in Table 6. The WVMD_DSN method shows improvements compared to single models GRU, CNN, and LSTM, with R^2 increasing by 3.7%, 4.2%, and 3.5%, respectively, and RMSE increasing by 7, 8, and 8 percentage points. MAE also increases by 5, 4.9, and 4.8 percentage points. Compared to the hybrid model CNN-BiLSTM, WVMD_DSN increases RMSE by around 6 percentage points, improves R^2 by 2.6%, and increases MAE by 3.7 percentage points. Similarly, compared to other hybrid models, WVMD_DSN shows corresponding improvements in RMSE, MAE, NRMSE, and R^2 . This validates that WVMD_DSN performs better than single models and hybrid models for wind power prediction. Compared to the original DSN, WVMD_DSN increases RMSE by approximately 5 percentage points, improves R^2 by 2.3%, and increases MAE by 2.9 percentage points. Therefore, the proposed WVMD_DSN model demonstrates superior predictive accuracy. The two streams of DSN proposed in this paper consist of SENet and BiGRU. The Table 6 shows that BiGRU has an RMSE 7.5% and 12.9% higher than GRU and LSTM, respectively. Thus, BiGRU is preferred for the second stream to handle wind power time series data. SENet, which is an improvement over CNN, has an RMSE 11.6% higher than CNN. Therefore, SENet is better for the first stream to capture spatial characteristics of wind power. An attention mechanism is then used to remove redundant data from both streams. Finally, combined with WVMD preprocessing technology, this approach shows a greater improvement compared to DSN alone. The prediction results are depicted in Fig. 7, showing substantial overlap between the WVMD_DSN approach and the original sequences, indicating its superior predictive performance.

Table 6: Experimental results of different models for wind power prediction

| Model | RMSE | MAE | NRMSE | R ² | EVS | MBE |
|-----------------|---------------|---------------|---------------|----------------|---------------|-------------|
| GRU [11] | 8.5975 | 5.6291 | 0.0426 | 0.9641 | 0.9647 | -1.13 |
| CNN [9] | 9.2461 | 5.4185 | 0.0458 | 0.9585 | 0.9634 | 3.15 |
| LSTM [8] | 9.1380 | 5.4973 | 0.0453 | 0.9595 | 0.9653 | 3.46 |
| BiGRU [14] | 7.9515 | 4.7374 | 0.0436 | 0.9693 | 0.9693 | -1.08 |
| CNN-LSTM [16] | 7.6317 | 4.7129 | 0.0378 | 0.9717 | 0.9737 | 2.05 |
| CNN-BiGRU [17] | 7.3976 | 4.6190 | 0.0366 | 0.9734 | 0.9737 | -0.76 |
| CNN-BiLSTM [18] | 7.2538 | 4.3032 | 0.0359 | 0.9744 | 0.9751 | 1.147 |
| DSN [21] | 6.8500 | 3.7651 | 0.0339 | 0.9772 | 0.9717 | -1.01 |
| WVMD_DSN | 1.0199 | 0.7298 | 0.0050 | 0.9995 | 0.9995 | 0.26 |

**Figure 7:** Prediction results and residual plot of WVMD_DSN

Compare the Combination of VMD with Multiple Models

The parameter settings for VMD and model parameters remain consistent with those previously described. VMD is combined with LSTM, CNN-LSTM, CNN-GRU, CNN-BiLSTM, and DSN, as shown in Table 7. These combinations significantly enhance prediction accuracy, with the DSN model demonstrating superior performance. Compared to LSTM, CNN-LSTM, CNN-GRU, and CNN-BiLSTM, integrating VMD with DSN results in RMSE increases of 65.5%, 26.9%, and 13.3%, respectively, MAE increases of 53%, 26.9%, and 2.7%, respectively, and R² increases of 4.4%, 1.3%, and 0.11%, respectively. This further confirms the DSN model's superiority in wind power prediction. Additionally, using WOA optimization further enhances prediction accuracy when combining VMD with DSN.

4.4.3 Generalization Analysis

In this section, the generalization ability of the WVMD_DSN model is primarily examined by using a wind power dataset 2. The optimized parameters (K , a) obtained from WVMD are (8, 100). Comparative evaluation metrics performance of WVMD_DSN and baseline models are shown in

Table 8. It can be observed that WVMD_DSN achieves an RMSE of 0.3704 and MAE of 0.2593. Compared to the single GRU model, these represent improvements of 89.6% and 88.9%, respectively. Compared to the hybrid CNN-BiGRU model, WVMD_DSN improves by 84.9% and 70.5%, and compared to the DSN model, improvements are 83.2% and 66.8%, respectively. The WVMD_DSN model continues to demonstrate superior performance on wind power dataset 2, highlighting its strong generalization capability.

Table 7: The combination of VMD and multiple models

| Methods | RMSE | MAE | R ² |
|---------------------|---------------|---------------|----------------|
| VMD-LSTM [34] | 7.7063 | 3.9861 | 0.9544 |
| VMD-CNN-GRU [35] | 5.6425 | 2.5621 | 0.9834 |
| VMD-CNN-BiLSTM [36] | 3.0625 | 1.9259 | 0.9954 |
| VMD-DSN | 2.6559 | 1.8733 | 0.9965 |
| WVMD_DSN | 1.0199 | 0.7298 | 0.9995 |

Table 8: Experimental results of different models for wind power dataset 2

| Model | RMSE | MAE | NRMSE | R ² |
|------------|---------------|---------------|---------------|----------------|
| GRU | 3.5829 | 2.3403 | 0.0808 | 0.9380 |
| CNN | 3.2300 | 2.1261 | 0.0654 | 0.9561 |
| LSTM | 3.1703 | 2.2935 | 0.0642 | 0.9577 |
| BiLSTM | 2.9219 | 1.9735 | 0.0592 | 0.9640 |
| CNN-GRU | 2.7328 | 1.7667 | 0.0553 | 0.9685 |
| CNN-LSTM | 2.7535 | 1.1669 | 0.0582 | 0.9601 |
| CNN-BiGRU | 2.4602 | 0.9832 | 0.0520 | 0.9682 |
| CNN-BiLSTM | 2.5754 | 1.2528 | 0.0544 | 0.9651 |
| DSN | 2.2087 | 0.7818 | 0.0508 | 0.9713 |
| WVMD_DSN | 0.3704 | 0.2593 | 0.0075 | 0.9994 |

5 Conclusion and Future Work

The paper proposes a short-term wind power prediction method that integrates WVMD with a novel DSN model. The main conclusions are as follows:

- (1) Through experiments on actual wind power data, it is shown that WVMD can address the adaptiveness in selecting the mode number K and penalty factor a more effectively.
- (2) The paper proposes a novel DSN model, which combines AM and SENet-BiGRU, demonstrating a significant improvement in the accuracy of wind power prediction. When combined with the WVMD preprocessing method, the effectiveness of wind power forecasting is further enhanced. Moreover, to verify the generalization capability of the WVMD_DSN method, experiments were conducted using real wind power dataset 2, which also exhibit volatility and instability, fully proving that WVMD_DSN has excellent generalization ability.

- (3) After a comprehensive analysis, the improved DSN model outperforms GRU, BiGRU, and CNN-BiGRU in both training time and prediction accuracy. While DSN shows better prediction accuracy, its training times increase by 20.4%, 36.1%, and 8.6% compared to GRU, BiGRU, and CNN-BiGRU, respectively. RMSE and MAE also increase by 21.1%, 14.6%, and 8.3% and 31.6%, 18.7%, and 16.6%, respectively. This model effectively enhances grid reliability and security and reduces scheduling time for integrating renewable energy into the power grid.

In the future, we will utilize some adaptive optimization algorithms to consider the learning rate and batch size parameters of the DSN network model. Additionally, we will use various data sources to improve the wind power generation forecasting model, including meteorological data, GIS data, wind farm layout data, and real-time monitoring data.

Acknowledgement: I am especially grateful to my teacher for her guidance and support in my research work. I also want to thank my laboratory colleagues for their help and suggestions throughout the research process.

Funding Statement: This work was supported in part by the Science and Technology Project of State Grid Corporation of China under Grant 5400-202117142A-0-0-00, the National Natural Science Foundation of China under Grant 62372242.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Yingnan Zhao, Yuyuan Ruan; data collection: Zhen Peng; analysis and interpretation of results: Yuyuan Ruan; draft manuscript preparation: Yuyuan Ruan, Yingnan Zhao. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in Wind-power-forecast at <https://github.com/ruanyuyuan/Wind-power-forecast.git> (accessed on 20 August 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. S. Ko, K. Lee, J. K. Kim, C. W. Hong, Z. Y. Dong and K. Hur, "Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 12, no. 2, pp. 1321–1335, 2021. doi: [10.1109/TSTE.2020.3043884](https://doi.org/10.1109/TSTE.2020.3043884).
- [2] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3276–3284, 2017. doi: [10.1109/TPWRS.2017.2787667](https://doi.org/10.1109/TPWRS.2017.2787667).
- [3] Z. Li *et al.*, "A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction," *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 39–54, 2023. doi: [10.1109/TSTE.2022.3198816](https://doi.org/10.1109/TSTE.2022.3198816).
- [4] P. Guo, Y. Gan, and D. Infield, "Wind turbine performance degradation monitoring using DPGMM and Mahalanobis distance," *Renew. Energy*, vol. 200, pp. 1–9, 2022. doi: [10.1016/j.renene.2022.09.115](https://doi.org/10.1016/j.renene.2022.09.115).
- [5] G. M. Shafiullah, M. T. Amanullah, A. B. Shawkat, P. Wolfs, and M. T. Arif, "Renewable energy integration: Opportunities and challenges," in *Ali. Smart Grids. Green Energy and Technology*. London: Springer, pp. 45–76, 2013.

- [6] J. Heineremann and O. Kramer, "Machine learning ensembles for wind power prediction," *Renew. Energy*, vol. 89, pp. 671–679, 2016. doi: [10.1016/j.renene.2015.11.073](https://doi.org/10.1016/j.renene.2015.11.073).
- [7] I. Khabbouchi, I. B. Salem, M. S. Guellouz, and U. Ritschel, "Machine learning and deep learning for wind power forecasting," in *2023 IEEE Int. Conf. Artif. Intell. Green Energy (ICAIGE)*, Sousse, Tunisia, 2023, pp. 1–6.
- [8] T. Ahmad and D. D. Zhang, "A data-driven deep sequence-to-sequence long-short memory method along with a gated recurrent neural network for wind power forecasting," *Energy*, vol. 239, 2022, Art. no. 122109. doi: [10.1016/j.energy.2021.122109](https://doi.org/10.1016/j.energy.2021.122109).
- [9] M. A. Hossain, R. K. Chakraborty, S. Elsayah, and M. J. Ryan, "Very short-term forecasting of wind power generation using hybrid deep learning model," *J. Clean. Prod.*, vol. 296, 2021, Art. no. 126564. doi: [10.1016/j.jclepro.2021.126564](https://doi.org/10.1016/j.jclepro.2021.126564).
- [10] Y. Y. Hong and C. Rioflorido, "A hybrid deep learning-based neural network for 24-h ahead wind power forecasting," *Appl. Energy*, vol. 250, pp. 530–539, 2019. doi: [10.1016/j.apenergy.2019.05.044](https://doi.org/10.1016/j.apenergy.2019.05.044).
- [11] P. Lu, L. Ye, Y. Zhao, B. Dai, M. Pei and Y. Tang, "Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges," *Appl. Energy*, vol. 301, 2021, Art. no. 117446. doi: [10.1016/j.apenergy.2021.117446](https://doi.org/10.1016/j.apenergy.2021.117446).
- [12] A. Meng *et al.*, "A novel few-shot learning approach for wind power prediction applying secondary evolutionary generative adversarial network," *Energy*, vol. 261, 2022, Art. no. 125276. doi: [10.1016/j.energy.2022.125276](https://doi.org/10.1016/j.energy.2022.125276).
- [13] A. A. Ezzat, M. Jun, and Y. Ding, "Spatio-temporal short-term wind forecast: A calibrated regime-switching method," *Ann. Appl. Stat.*, vol. 13, no. 3, pp. 1484–1510, 2019. doi: [10.1214/19-AOAS1243](https://doi.org/10.1214/19-AOAS1243).
- [14] A. A. Ezzat, M. Jun, and Y. Ding, "Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1437–1447, 2018. doi: [10.1109/TSTE.2018.2789685](https://doi.org/10.1109/TSTE.2018.2789685).
- [15] S. Hu *et al.*, "Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction," *Appl. Energy*, vol. 293, 2021, Art. no. 116951. doi: [10.1016/j.apenergy.2021.116951](https://doi.org/10.1016/j.apenergy.2021.116951).
- [16] J. Zhang *et al.*, "Power prediction of a wind farm cluster based on spatiotemporal correlations," *Appl. Energy*, vol. 302, 2021, Art. no. 117568. doi: [10.1016/j.apenergy.2021.117568](https://doi.org/10.1016/j.apenergy.2021.117568).
- [17] Y. Chen *et al.*, "CNN-BiLSTM short-term wind power forecasting method based on feature selection," *IEEE J. Radio Freq. Identif.*, vol. 6, pp. 922–927, 2022. doi: [10.1109/JRFID.2022.3213753](https://doi.org/10.1109/JRFID.2022.3213753).
- [18] Z. Ma and G. Mei, "A hybrid attention-based deep learning approach for wind power prediction," *Appl. Energy*, vol. 323, 2023, Art. no. 119608. doi: [10.1016/j.apenergy.2022.119608](https://doi.org/10.1016/j.apenergy.2022.119608).
- [19] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," In: *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, Curran Associates, Inc., 2014, vol. 27, pp. 568–576.
- [20] B. Su, P. Zhang, M. Sun, and M. Sheng, "Direction-guided two-stream convolutional neural networks for skeleton-based action recognition," *Soft Comput.*, vol. 27, pp. 11833–11842, 2023. doi: [10.1007/s00500-023-07862-1](https://doi.org/10.1007/s00500-023-07862-1).
- [21] Z. A. Khan, T. Hussain, and S. W. Baik, "Dual stream network with attention mechanism for photovoltaic power forecasting," *Appl. Energy*, vol. 338, 2023, Art. no. 120916. doi: [10.1016/j.apenergy.2023.120916](https://doi.org/10.1016/j.apenergy.2023.120916).
- [22] W. Zhang, F. Liu, X. Zheng, and Y. Li, "A hybrid EMD-SVM based short-term wind power forecasting model," in *2015 IEEE PES Asia-Pacific Power Energy Eng. Conf. (APPEEC)*, Brisbane, QLD, Australia, 2015, pp. 1–5.
- [23] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method," *Renew. Energy*, vol. 94, pp. 629–636, 2016. doi: [10.1016/j.renene.2016.03.103](https://doi.org/10.1016/j.renene.2016.03.103).
- [24] I. Daubechies, "The wavelet transform time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, 1990. doi: [10.1109/18.57199](https://doi.org/10.1109/18.57199).

- [25] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 1, 2014. doi: [10.1109/TSP.2013.2288675](https://doi.org/10.1109/TSP.2013.2288675).
- [26] N. Rehman and D. P. Mandic, "Empirical mode decomposition for trivariate signals," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1059–1068, 2010. doi: [10.1109/TSP.2009.2033730](https://doi.org/10.1109/TSP.2009.2033730).
- [27] X. Gao, W. Guo, C. Mei, J. Sha, Y. Guo and H. Sun, "Short-term wind power forecasting based on SSA-VMD-LSTM," *Energy Rep.*, vol. 9, pp. 335–344, 2023. doi: [10.1016/j.egy.2023.05.181](https://doi.org/10.1016/j.egy.2023.05.181).
- [28] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016. doi: [10.1016/j.advengsoft.2016.01.008](https://doi.org/10.1016/j.advengsoft.2016.01.008).
- [29] D. Wang, D. Tang, and L. Liu, "Particle swarm optimization algorithm," *Soft Comput.*, vol. 22, pp. 387–408, 2018. doi: [10.1007/s00500-016-2474-6](https://doi.org/10.1007/s00500-016-2474-6).
- [30] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014. doi: [10.1016/j.advengsoft.2013.12.007](https://doi.org/10.1016/j.advengsoft.2013.12.007).
- [31] N. Kwak and C. H. Choi, "Input feature selection for classification problems," *IEEE Trans. Neural Netw.*, vol. 13, pp. 143–159, 2002.
- [32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 7132–7141.
- [33] K. Cho, B. Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *8th Workshop Syntax, Semant. Struct. Stat. Transl., SSST 2014*, Doha, Qatar, 2014, pp. 103–111. doi: [10.3115/v1/W14-40](https://doi.org/10.3115/v1/W14-40).
- [34] L. Han, R. Zhang, X. Wang, A. Bao, and H. Jing, "Multi-step wind power forecast based on VMD-LSTM," *IET Renew. Power Gener.*, vol. 13, pp. 1609–1700, 2019. doi: [10.1049/iet-rpg.2018.5781](https://doi.org/10.1049/iet-rpg.2018.5781).
- [35] Z. Zhao *et al.*, "Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features," *Eng. Appl. Artif. Intell.*, vol. 121, 2023, Art. no. 105982. doi: [10.1016/j.engappai.2023.105982](https://doi.org/10.1016/j.engappai.2023.105982).
- [36] Y. Dai, R. Wang, Y. Ma, T. Wan, and Z. Huang, "Research on CNN-BiLSTM power load forecasting based on VMD algorithm," in *2023 IEEE 5th Int. Conf. Civil Aviation Saf. Inform. Technol. (ICCASIT)*, Dali, China, 2023, pp. 1098–1102.