

A SHORT-TERM WIND POWER FORECASTING METHOD BASED ON HOA-VMD-DWE

YANG, Y. C.[#] – SUN, D. H.[#] – SUN, J. J.^{*}

School of Electrical Engineering and Automation, Wuhan University, Wuhan, Hubei 430072, China

[#]These authors contributed equally to this work and should be considered co-first authors

^{}Corresponding author
e-mail: jjsun@whu.edu.cn*

(Received 8th Oct 2025; accepted 30th Jan 2026)

Abstract. Wind power generation exhibits prominent volatility, nonlinearity, and randomness, which pose significant challenges to the safe and stable operation of power systems. To improve the accuracy of wind power prediction, this paper proposes a wind power forecasting model (HOA-VMD-DWE) that integrates the Hiking Optimization Algorithm (HOA), the Variational Mode Decomposition (VMD), and the Dynamic Weighted Ensemble (DWE). First, HOA is employed to globally optimize the key parameters of the VMD, thereby enhancing the quality of modal decomposition. Subsequently, VMD decomposes the original wind power sequence to improve the time-frequency expression ability of features. On this basis, the Gated Recurrent Unit (GRU) and the Multi-Layer Perceptron (MLP) are respectively used to predict meteorological features and decomposed IMF sequences. Finally, a DWE network is introduced to dynamically fuse the prediction results of multiple models, aiming to improve prediction accuracy and robustness. Experimental results demonstrate that the proposed model achieves superior predictive performance under different operating conditions, including low wind, strong gusts, and general wind conditions, and can provide effective support for the stable operation and dispatch decision-making of power grids.

Keywords: *nonlinear time-series modeling, multi-scale feature decomposition, swarm intelligence algorithm, deep ensemble learning, complex meteorological conditions, power sequence prediction*

Introduction

In recent years, wind power has become an essential component of renewable energy, with installed capacity increasing significantly worldwide. However, wind power output is subject to strong uncertainty and volatility due to meteorological and environmental influences, posing considerable challenges to the efficiency of power system dispatch and the stability of power system operation. Therefore, achieving highly accurate and generalizable wind power forecasting is of great importance for improving wind energy utilization efficiency and ensuring the safe and stable operation of power grids. This research direction holds both significant theoretical value and engineering application prospects.

At present, a wide variety of wind power forecasting methods have been proposed. Based on the mathematical modeling approach employed, these methods can be broadly categorized into four groups: physical models, statistical models, machine learning models, and hybrid models. Physical models construct predictive frameworks by incorporating environmental data and power curves, which can improve forecasting accuracy. Nevertheless, these models require high-quality data and extensive domain knowledge, while their robustness is limited, making large-scale applications difficult. Statistical models establish functional relationships between meteorological variables and power output by analyzing power and weather data. For example, Zhou et al. (2011)

proposed a short-term forecasting method based on statistical clustering, which better captures the randomness and uncertainty of wind power and is suitable for short-term prediction. However, such models depend heavily on clustering algorithms and data quality, leading to limited generalization and unsatisfactory performance in complex wind farms or long-term forecasting. Similarly, Stathopoulos et al. (2013) combined numerical and statistical models to achieve more comprehensive prediction, improving accuracy and stability, but at the cost of increased model complexity, higher computational requirements, and limited scalability.

Machine learning approaches, represented by long short-term memory (LSTM) networks and random forests, have attracted significant attention in recent years. For instance, Zhou et al. (2016) applied a random forest method, which can handle high-dimensional data with strong robustness and resistance to overfitting. Nevertheless, its interpretability is limited, and its accuracy may decline when dealing with strong nonlinearities and temporal dependencies. Xu et al. (2025) proposed that a cross-dataset neural network benchmark was developed to better evaluate generalization ability across different wind farm scenarios, though it requires extensive data, incurs high computational costs, and is sensitive to missing data and noise. Hybrid models integrate the advantages of different methods, capturing temporal dependencies while reducing noise through decomposition or filtering. These methods often outperform traditional statistical models, but they demand large amounts of high-quality data, have high computational complexity, and exhibit limited interpretability, which hinders engineering applications. For example, Liu et al. (2019) combined LSTM with wavelet transform for multi-scale feature extraction; Huang et al. (2024) employed EEMD-LSTM to improve noise resistance in ultra-short-term forecasting; Ahmadi et al. (2025) integrated Kalman filtering, deep residual networks, and bidirectional LSTM into a comprehensive hybrid framework; Khan et al. (2025) combined LSTM-SMI with ARIMA, leveraging both statistical and deep learning strengths; while Li et al. (2013) proposed a CFD-based physical flow field method, which offers strong interpretability and physical fidelity but suffers from prohibitive computational costs that limit real-time application.

In response to the ongoing challenges in wind power forecasting, recent studies have further explored the combination of signal decomposition, metaheuristic optimization, and advanced learning architectures. For instance, You et al. (2024) employed an enhanced VMD model optimized by the dung beetle algorithm with a kernel extreme learning machine to improve prediction accuracy. Similarly, Hou et al. (2024) proposed a multistep forecasting framework based on secondary decomposition and an enhanced arithmetic optimization algorithm. Other hybrid approaches include the use of improved temporal convolutional networks (Xin et al., 2025), adaptive VMD with deep learning autoencoders (Rayi et al., 2021), dual-optimization models integrating deep learning and bio-inspired algorithms (Li et al., 2024), as well as IHOA-based wind speed prediction (Lin et al., 2025). Moreover, ensemble and decomposition-enhanced methods such as VMD with bagging extreme learning (Ribeiro et al., 2024), VMD-SSA-TCN-BiGRU hybrids (Zhang et al., 2023), hybrid optimization with wind load data (Qin et al., 2021), and hierarchical models with optimized feature decomposition (Liu et al., 2025) have also demonstrated promising performance in handling the non-stationarity and complexity of wind power series. These works collectively highlight the importance of integrating adaptive decomposition, parameter optimization, and flexible modeling structures, which aligns with the motivation of the present study.

In this context, this paper proposes a short-term wind power forecasting approach based on Variational Mode Decomposition–Dynamic Weighted Ensemble (VMD-DWE), optimized by the Hiking Optimization Algorithm (HOA). The HOA is employed to perform global optimization of the core VMD parameters, effectively avoiding the local optima problem commonly encountered in traditional gradient-based optimization methods. The VMD algorithm is then used to decompose the original wind power sequence into intrinsic mode functions (IMFs) and a residual sequence, each capturing distinct temporal scale characteristics. These IMFs and the residuals are further integrated with meteorological input features, including wind speed, wind direction, humidity, temperature, and air pressure, to construct a comprehensive feature set containing multi-source heterogeneous information. This enhances the hierarchical structure and time-frequency resolution of the feature representation, thereby facilitating the exploration of intrinsic driving mechanisms underlying wind power fluctuations.

Compared with traditional single-model approaches, the proposed method first applies gated recurrent units (GRU) and multilayer perceptrons (MLP) independently to forecast wind power, and subsequently develops a dynamic weighted ensemble (DWE) framework. The DWE adaptively adjusts the fusion weights of GRU and MLP according to the input features, thereby improving overall prediction accuracy, accelerating convergence, and enhancing robustness against high-frequency disturbances. This method provides a reliable technical foundation for high-precision wind power forecasting.

Model construction

HOA model

The Hiking Optimization Algorithm (HOA) is a swarm intelligence optimization method inspired by human hiking and exploration behavior (Zhang et al., 2022). It simulates the process by which hikers search for optimal climbing routes in complex terrain, particularly the “exploration–review–collaboration” strategy, to guide individuals in dynamically searching the solution space and approaching the global optimum. First, HOA initializes several “hiker” agents randomly in the solution space and evaluates their fitness to assess the quality of their current positions. Subsequently, three key behavioral strategies are introduced to achieve global optimization. (1) Local climbing: hikers ascend according to the current terrain, thereby improving the local search precision of individuals. (2) Path review: each agent dynamically adjusts its direction based on historical trajectories and current fitness feedback, preventing oscillations or the continuation of incorrect paths. (3) Group collaborative navigation: information sharing among individuals drives the population toward the optimal region, thereby improving overall search efficiency and global convergence capability. Finally, the algorithm iteratively updates the positions and fitness evaluations of the agents until the convergence criteria are satisfied (e.g., reaching the maximum number of iterations or achieving a predefined fitness threshold). The theoretical derivation is as follows:

(1) Population initialization

Within the parameter search space, initial hiker positions are randomly generated, where each position corresponds to a candidate set of VMD parameters:

$$X = \{x_1, x_2, \dots, x_N\}, x_i = (\alpha_i, K_i) \quad (\text{Eq.1})$$

where N denotes the population size, while $\alpha_i \in [\alpha_{\min}, \alpha_{\max}]$ and $K_i \in [K_{\min}, K_{\max}]$ represent the search ranges of the penalty factor and the number of decomposition modes, respectively.

(2) Explorer exploration

Explorers expand new search paths based on gradient information, with position updates defined as:

$$x_{i,j}^{t+1} = x_{i,j}^t + \gamma \cdot \nabla f(x_{i,j}^t) \cdot \text{rand}() \quad (\text{Eq.2})$$

where γ is the exploration step size, ∇f is the gradient of the objective function, and $\text{rand}()$ is a uniformly distributed random number.

(3) Follower learning

Followers learn from the current optimal individual while maintaining population diversity:

$$x_{i,j}^{t+1} = x_{\text{best},j}^t + \beta \cdot |x_{i,j}^t - x_{\text{best},j}^t| \cdot \text{randn}() \quad (\text{Eq.3})$$

where x_{best} is the current optimal position, β is the learning factor, and $\text{randn}()$ denotes a standard normally distributed random variable.

(4) Sentinel perturbation

Sentinels introduce stochastic jumps to prevent the algorithm from being trapped in local optima:

$$x_{i,j}^{t+1} = x_{\min,j} + (x_{\max,j} - x_{\min,j}) \cdot \text{rand}() \quad (\text{Eq.4})$$

where $x_{\max,j}$ and $x_{\min,j}$ denote the upper and lower bounds of the j th parameter, respectively.

(5) Basic procedure

Figure 1 illustrates the detailed procedural framework of the HOA algorithm to clarify its execution logic.

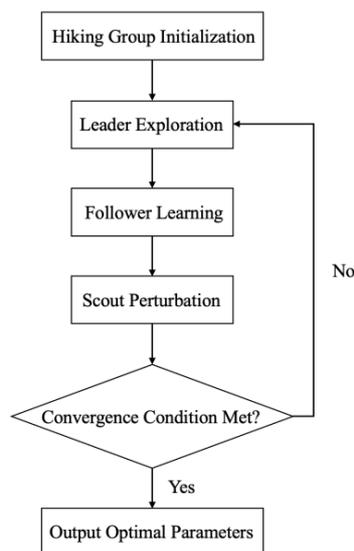


Figure 1. Flowchart of the HOA algorithm

- ① Initialization: Randomly generate hiker positions (VMD parameter combinations) and compute the objective function values (a combination of reconstruction error and correlation penalty);
- ② Iterative update: Partition individuals into explorers, followers, and sentinels according to their performance, and execute corresponding exploration, learning, and perturbation operations;
- ③ Termination: Output the optimal VMD parameters (α^*, K^*) once the maximum iteration count is reached or convergence of the objective function is achieved.

VMD model

VMD is an adaptive signal decomposition method grounded in variational theory and signal processing optimization, designed to address the mode-mixing issue encountered by traditional Empirical Mode Decomposition (EMD) when processing nonlinear signals (Duan, 2025). VMD constructs a variational optimization model that decomposes a signal into several Intrinsic Mode Functions (IMFs) with limited bandwidth and concentrated spectra, where the center frequency of each mode is adaptively updated during iterations. The algorithm minimizes the bandwidth of each mode under the constraint of signal reconstruction by introducing a penalty factor, reformulating the problem into an unconstrained variational optimization solved efficiently via the Alternating Direction Method of Multipliers (ADMM). The decomposition performance of VMD is influenced primarily by two parameters: the number of modes K , which determines the resolution and granularity of decomposition, and the penalty factor, which controls the bandwidth of each mode and tunes its sensitivity to noise and fine details. By adaptively optimizing these parameters according to the characteristics of the target signal, VMD achieves higher decomposition precision and stronger noise robustness, enabling optimal performance across diverse data types and application scenarios. The core derivation is as follows:

(1) Constrained variational model

The objective is to minimize the squared norm of each mode, subject to the constraint that the sum of all modes equals the original signal:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (\text{Eq.5})$$

$$\text{s.t. } \sum_{k=1}^K u_k = f \quad (\text{Eq.6})$$

where u_k is the k th mode, ω_k its center frequency, $\delta(t)$ the Dirac delta function, $*$ the convolution operator, and f the original wind power signal.

(2) Iterative update procedure

① Mode function update:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (\text{Eq.7})$$

② Center frequency update:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (\text{Eq.8})$$

③ Lagrange multiplier update:

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega) \right) \quad (\text{Eq.9})$$

(3) Basic procedure

Figure 2 illustrates the detailed procedural framework of the VMD algorithm to clarify its execution logic.

- ① Initialize mode functions, center frequencies, and Lagrange multipliers;
- ② Iteratively update all parameters using the above equations until the variation in each mode is smaller than a predefined threshold;
- ③ Output the decomposed K IMFs.

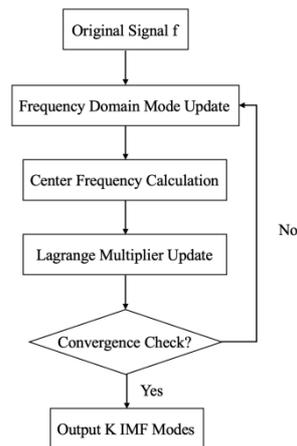


Figure 2. Flowchart of the VMD algorithm

Neural network models

GRU model

The Gated Recurrent Unit (GRU) captures the temporal dependencies of wind power data through reset and update gates (Wang et al., 2022). The computational process is defined as follows:

(1) Reset gate – controls the degree of forgetting of historical information:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (\text{Eq.10})$$

(2) Update gate – balances the weights of new and historical information:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (\text{Eq.11})$$

(3) Candidate state – generates new candidate memory:

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \quad (\text{Eq.12})$$

(4) Final state – integrates historical and current information:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (\text{Eq.13})$$

where σ denotes the sigmoid activation function, \odot the element-wise product, h_t the hidden state at time t , W_r, W_z, W_h the weight matrices, and b_r, b_z, b_h the bias terms.

Figure 3 depicts the operational flow of the GRU model.

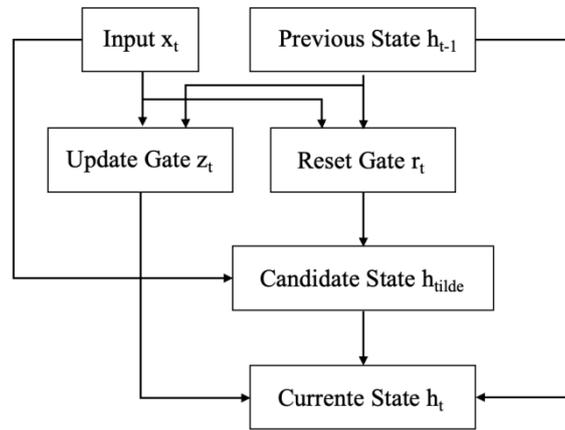


Figure 3. Flowchart of the GRU algorithm

MLP model

The Multilayer Perceptron (MLP) learns the nonlinear mapping characteristics of wind power through multiple layers of nonlinear transformations (Sireesha and Thotakura, 2024). The forward propagation is expressed as:

$$y = f_L(W_L \cdot f_{L-1}(\dots f_1(W_1 \cdot x + b_1) \dots) + b_L) \quad (\text{Eq.14})$$

where f_i denotes the ReLU activation function ($f(x) = \max(0, x)$), and W_i represent the weights and b_i the biases of the i th layer, respectively, and L is the number of network layers.

DWE model

The core principle of the Dynamic Weight Ensemble (DWE) network lies in adaptively integrating multi-model prediction outputs through a dynamic weight allocation mechanism based on the softmax function, thereby improving overall forecasting performance (Yu and Dai, 2022). In this study, the number of sub-models N is set to two, corresponding to a Gated Recurrent Unit (GRU) network and a Multilayer Perceptron (MLP). Specifically, the process involves feeding the input data into GRU and MLP models for independent predictions, dynamically generating weight coefficients using the softmax function according to their performance on the current task, and then fusing the outputs through weighted averaging to yield the final ensemble prediction. The dynamic weight adjustment mechanism of DWE effectively avoids the performance degradation associated with fixed-weight strategies when the data distribution changes. By combining the temporal modeling strength of GRU with the nonlinear mapping capability of MLP, DWE significantly enhances forecasting accuracy and generalization capability. The core process is as follows:

(1) Weight calculation

Normalized weights are generated from the hidden state:

$$w_t = \text{softmax}(W_w \cdot \text{ReLU}(W_h \cdot h_t + b_h) + b_w) \quad (\text{Eq.15})$$

with softmax ensuring that the sum of weights equals 1:

$$w_{i,t} = \frac{e^{z_{i,t}}}{\sum_{j=1}^2 e^{z_{j,t}}} \quad (\text{Eq.16})$$

where h_t is the GRU hidden state, W_w, W_h the weight matrix, and b_w, b_h the bias term.

(2) Ensemble prediction

The final prediction result is obtained as a weighted sum:

$$\hat{y}_t = \sum_{i=1}^2 w_{i,t} \cdot \hat{y}_{i,t} \quad (\text{Eq.17})$$

where $\hat{y}_{1,t}$ and $\hat{y}_{2,t}$ denote the predictions of GRU and MLP, respectively.

Figure 4 illustrates the execution flow of the DWE algorithm.

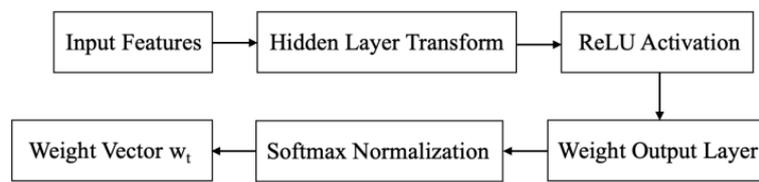


Figure 4. Flowchart of the DWE algorithm

Experimental design

Dataset partitioning and preprocessing

To ensure the rigor of the wind power prediction experiment in terms of temporal causality, this study adopts a chronological splitting strategy based on the timeline to partition the original wind power data and the corresponding multivariate meteorological data.

Specifically, the complete time series is first divided into three parts: the training set, the validation set, and the test set in chronological order. Each dataset is continuous and non-overlapping in the time dimension. Among them, the training set is utilized for the optimization of VMD parameters within the HOA-VMD-DWE model and the learning of sub-model parameters. The validation set is employed for model hyperparameter selection and early stopping control during the training process. The test set is reserved exclusively for the final evaluation of the model's generalization performance.

During the dataset partitioning stage, sliding window samples are not directly constructed; instead, the temporal boundary division for the training, validation, and test sets is completed solely at the level of the original time series. The construction of sliding window samples and the rolling prediction operations are integrated into the model prediction process. This approach ensures a clear separation between the dataset partitioning and the model execution mechanism in terms of temporal logic.

Prediction process of the HOA-VMD-DWE model

Upon completion of the dataset partitioning, the overall prediction workflow of the HOA-VMD-DWE model is constructed based on a sliding time window mechanism. The specific prediction procedure is as follows:

1. Data Preprocessing: Standardizing the original wind power and meteorological data to eliminate dimensional discrepancies and performing missing value imputation.

2. Implementation of the Sliding Time Window Method is introduced to organize the wind power prediction process. The sampling interval for the wind power and meteorological data in this study is 15 min. Considering the practical requirements of ultra-short-term wind farm operation scheduling and power smoothing control, the prediction time scale is primarily focused on approximately 15 min. Based on this engineering context, the prediction step size is set to $H = 1$, corresponding to a 15-min ahead wind power prediction. This ensures the practical guiding significance of the prediction results while avoiding the issue of amplified uncertainty associated with excessively long prediction steps. Furthermore, considering that wind speed and wind power series typically exhibit significant temporal correlation within a scale of several hours, and that the HOA-optimized Variational Mode Decomposition (VMD) can decompose non-stationary power signals into several narrow-band sub-sequences, the dynamic characteristics of each sub-sequence remain relatively concentrated. This reduces the dependency on long-term historical information. To balance the characterization of wind condition evolution with model complexity, the sliding window length is set to $L = 32$, which corresponds to 8 h of historical data.

At any prediction instant t , the model exclusively leverages the historical wind power time series and corresponding multi-variable meteorological observations within the sliding window $[t-L + 1, t]$ as inputs, with the prediction target being the wind power output at the immediate next time step $t + 1$. This sliding-window paradigm permeates the entire HOA-VMD-DWE framework, where HOA-based parameter optimization, VMD decomposition, feature engineering, sub-model inference, and dynamic ensemble are executed independently for each window.

3. The historical wind power series contained in the current window is fed into the HOA optimization module. By minimizing a holistic objective function that incorporates both reconstruction error and an IMF cross-correlation penalty term, the key hyperparameters of VMD are optimized, yielding the optimal configuration (α^*, K^*) .

4. The optimized parameters are then employed to perform VMD decomposition on the windowed wind power series, decomposing it into K intrinsic mode functions (IMFs) with distinct frequency signatures. Since the decomposition is constrained strictly to historical data within the current window, the resulting IMFs are derived in a temporally causal manner from observed measurements, providing a robust multi-scale representation to underpin subsequent predictive modeling.

5. During the feature engineering phase, rigorous temporal alignment is applied to each decomposed IMF: specifically, the IMF subsequence and its corresponding multi-variable meteorological features at the same time index are concatenated to form a joint input feature vector for the respective IMF channel.

6. In the sub-model inference phase, the input features of each IMF channel are routed in parallel to two heterogeneous sub-models—GRU and MLP. The GRU specializes in capturing temporal dependencies within the sliding window, thereby extracting short- to medium-term dynamic patterns from the IMF components. In contrast, the MLP focuses on modeling the nonlinear mapping between IMF components and meteorological

covariates. Both sub-models are trained using the Adam optimizer to accelerate convergence and enhance prediction robustness.

7. In the dynamic ensemble phase, a Deep Weighted Ensemble (DWE) network is introduced to adaptively fuse the outputs of the GRU and MLP. The DWE computes fusion weights dynamically based on the hidden states or feature embeddings of the sub-models within the current window, enabling adaptive balancing of the contributions from temporal modeling and nonlinear fitting across varying time scales and operational scenarios. The dynamically weighted predictions from each IMF channel are then aggregated and reconstructed to produce the final wind power forecast for the target time instant.

Through the above rolling prediction process based on a sliding time window, the HOA-VMD-DWE model adheres strictly to temporal causality constraints throughout the prediction procedure. This not only effectively mitigates the issue of future information leakage introduced by signal decomposition methods, but also ensures that the model's operational mechanism is highly consistent with the practical scenarios of online wind power prediction, thereby laying a solid foundation for subsequent experimental analysis and performance evaluation. The overall architecture of the model is illustrated in *Figure 5*.

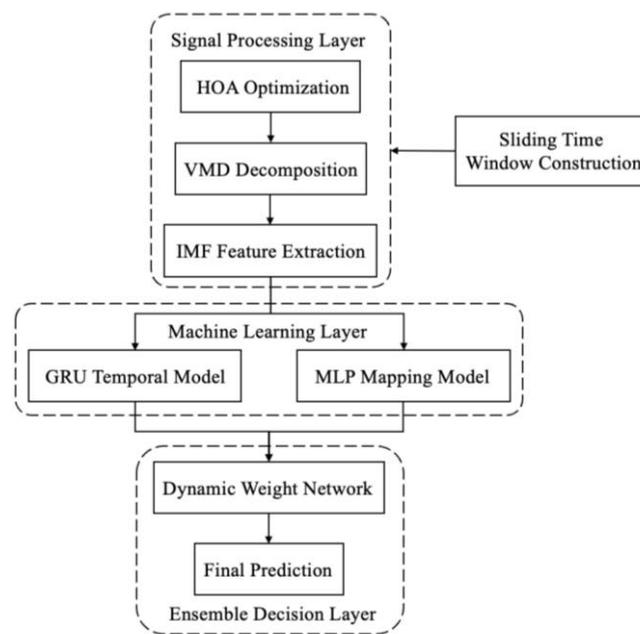


Figure 5. Prediction flow chart of the HOA-VMD-DWE framework

Selection of experimental analysis samples

Following data cleaning and the design of the modeling workflow, a systematic statistical analysis and type screening of the raw data were conducted to comprehensively and objectively evaluate the applicability and robustness of the proposed prediction model under varying wind conditions.

Based on the full-year 15-min resolution wind power and multi-source meteorological data from the Guohua Jingxia Wind Farm with an installed capacity of 200 MW in Xinjiang, China, in 2019, wind conditions were classified and categorized from multiple

statistical perspectives, including wind speed distribution characteristics, fluctuation intensity, and abrupt change behavior.

Statistical results indicate that three dominant operating scenarios—low wind speed, strong gust disturbance, and general wind conditions—prevail throughout the year, covering the typical meteorological conditions and power variation patterns encountered in the actual operation of wind power generation systems. Therefore, selecting these three representative weather scenarios as experimental analysis samples enables a relatively comprehensive validation of the model's predictive performance under different operating conditions.

1. Under low wind conditions, the wind power signal exhibits low amplitude and slow variation, which imposes high requirements on the prediction model's ability to capture long-term trends and subtle features. Thus, this sample type is primarily used to validate the model's capability to fit power change trends under low signal-to-noise ratio conditions and its prediction stability for minor fluctuations. The selection criteria are wind speed is less than the 20th percentile of the historical distribution (3.20 m/s) for more than 80% of the day, and the daily wind speed standard deviation is less than the threshold σ_2 (overall standard deviation $4.73 \text{ m/s} \times 0.4 = 1.89 \text{ m/s}$).

2. Under transient strong gust conditions, the wind power sequence demonstrates strong non-stationarity and abruptness, which demands higher dynamic responsiveness, short-term prediction accuracy, and robustness to sudden events from the model. Therefore, this sample type is mainly employed to evaluate the model's ability to characterize wind power ramps and drop-off processes under high-fluctuation, rapid-change conditions, with a focus on testing its capture of complex temporal features. The selection criteria are: within a 30-min time window, the wind speed increases abruptly by $\geq 30\%$, and the maximum wind speed reaches or exceeds the 80th percentile of the historical distribution.

3. General wind conditions account for the largest proportion of annual operation and reflect the typical operational status of the wind farm under normal meteorological conditions. As such, this sample type is used to comprehensively assess the overall prediction accuracy and stability of the model under routine operating conditions, reflecting its average performance level in engineering applications. The selection criteria are: wind speed lies within the 20th–80th percentile interval of the historical distribution (3.20 – 11.99 m/s) for no less than 70% of the day; there are no events where wind speed increases abruptly by $\geq 30\%$ within 30 min and peaks at ≥ 80 th percentile throughout the day; and the daily wind speed standard deviation is less than the threshold σ_2 (overall standard deviation $4.73 \text{ m/s} \times 0.4 = 1.89 \text{ m/s}$).

Results and discussion

Model performance evaluation

To comprehensively and objectively evaluate the accuracy and reliability of the proposed wind power forecasting model, four error metrics are employed: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2).

The evaluation metrics are defined as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{Eq.18})$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{Eq.19})$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{Eq.20})$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{Eq.21})$$

where y_i denotes the predicted value, n is the total number of samples, and \bar{y} represents the mean of the actual values.

Model forecasting

Forecasting under typical weak-wind conditions

Under typical weak-wind conditions, the amplitude of wind power output variations is relatively small, and the output curve exhibits smooth fluctuations. After processing the dataset corresponding to weak-wind conditions and feeding it into the forecasting models, the predicted values and error curves of different models are shown in *Figures 6 and 7*, respectively, while the comparative performance evaluation results are summarized in *Table 1*.

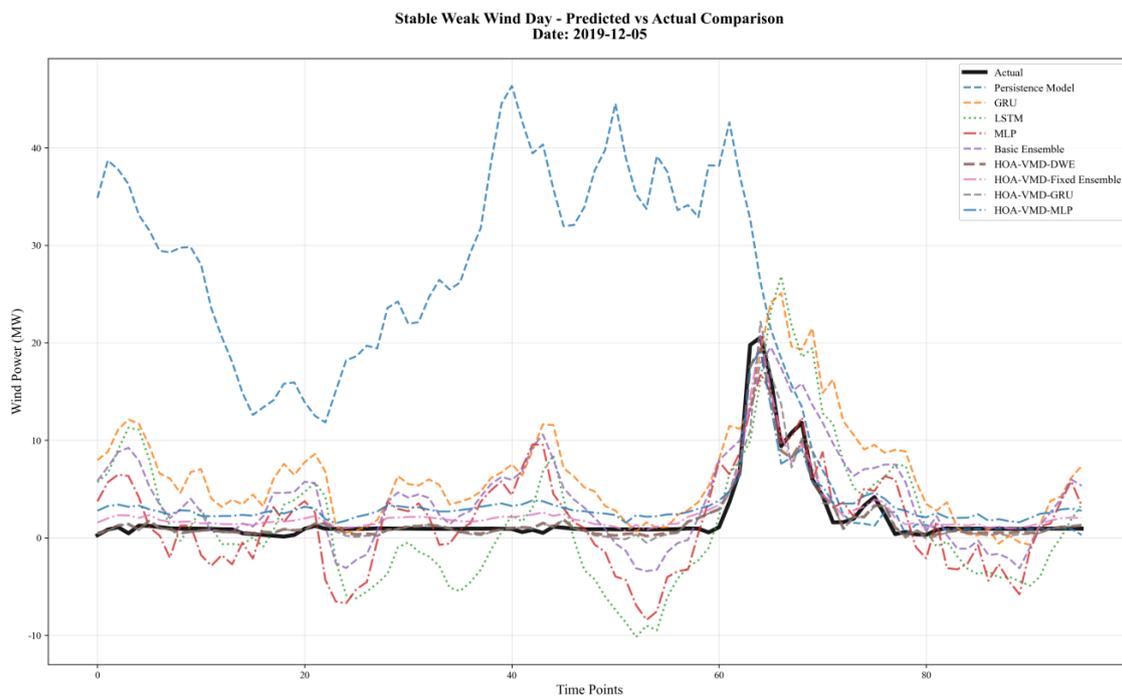


Figure 6. Prediction vs. actual values under weak-wind conditions

As shown in *Figures 6 and 7*, the single models (GRU, MLP, and LSTM) as well as the baseline ensemble model present noticeable deviations from the actual measurements, leading to relatively low fitting accuracy. The day-ahead 24-h persistence forecasting method similarly exhibits significant prediction deficiencies. Its results show notable

deviations in tracking actual wind power fluctuations during power variation processes, making it difficult to effectively characterize subtle variations and local abrupt changes under low wind speed conditions. In contrast, the four HOA-VMD-optimized models exhibit superior forecasting performance. Among them, the HOA-VMD-DWE model achieves the lowest prediction errors, with its error curve most closely aligned with the zero-error line, highlighting its significant advantage in prediction accuracy under weak-wind scenarios.

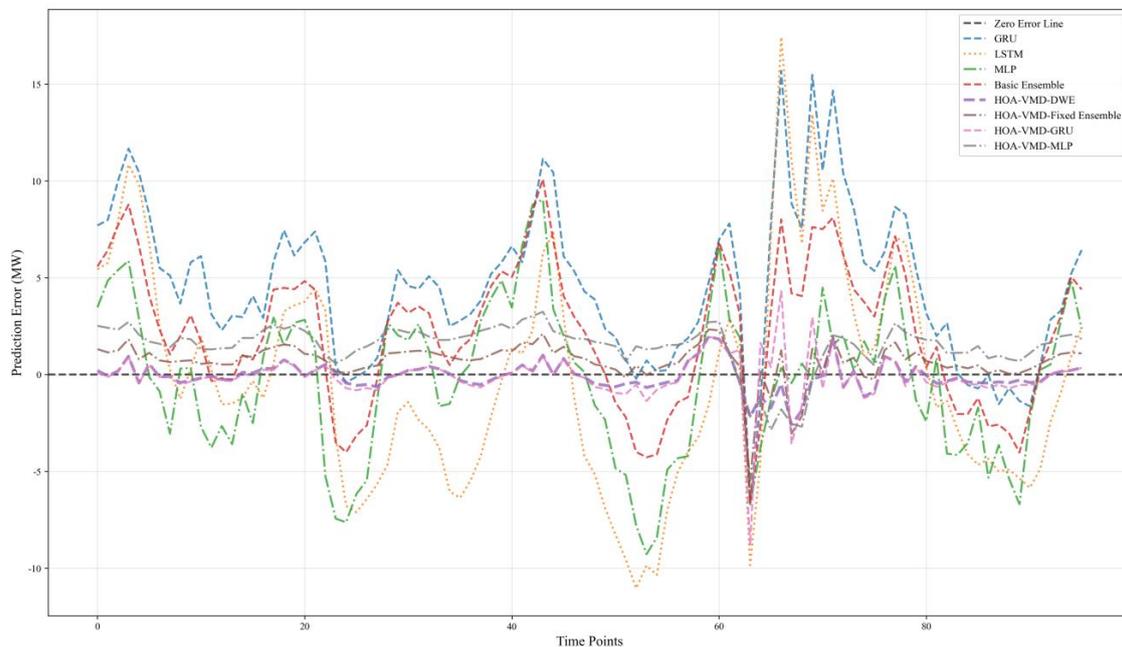


Figure 7. Prediction error curves under weak-wind conditions

Based on the analysis of the results in *Figure 8*, the DWE mechanism exhibits distinct adaptive adjustment characteristics during the forecasting process. During periods of weak wind and gentle variation, the weights of the GRU model are maintained at a high level for extended durations, indicating its stable advantage in characterizing the temporal continuity and inertial features of wind power sequences under low wind speed conditions. However, comparative analysis reveals that during certain phases of relatively stronger wind within stable weak wind weather, the weight proportion of the MLP model shows a significant increase. This phenomenon indicates that when wind speed increases briefly and power variations exhibit stronger magnitude-dominated characteristics rather than obvious temporal dependencies, the MLP model possesses a superior capability in characterizing instantaneous nonlinear mapping relationships. Consequently, it is adaptively assigned higher weights by the DWE mechanism during the ensemble process. The aforementioned results further validate that the DWE mechanism can flexibly allocate weights among different models according to the dynamic characteristics of wind power. This enables an effective response to local structural changes under weak wind conditions, thereby enhancing the robustness and accuracy of the overall prediction.

According to *Table 1*, the optimized models—HOA-VMD-GRU, HOA-VMD-Fixed Ensemble, HOA-VMD-MLP, and HOA-VMD-DWE—consistently outperform the GRU, MLP, and LSTM single models as well as the baseline ensemble model across

all error metrics (MSE, MAE, RMSE). In particular, the HOA-VMD-DWE model achieves the best overall performance, with MSE, MAE, and RMSE values of 0.5087, 0.5080, and 0.7133, respectively—the lowest among all models. Moreover, its coefficient of determination (R^2) reaches 0.9612, closest to 1, demonstrating its capability to capture wind power output trends with high forecasting precision under weak-wind conditions.

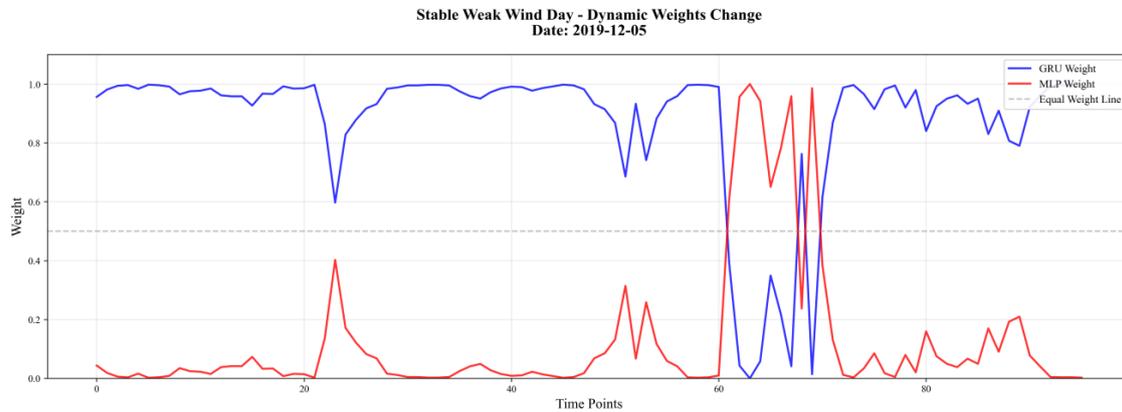


Figure 8. DWE weight curve under weak-wind conditions

Table 1. Error evaluation metrics of different models under weak-wind conditions

Model	MSE	MAE	RMSE	R^2
GRU	36.4358	4.8981	6.0362	-1.7756
LSTM	31.6033	4.5465	5.6217	-1.4074
Basic ensemble	17.6288	3.5277	4.1987	-0.3429
MLP	15.3531	3.1707	3.9183	-0.1696
HOA-VMD-MLP	3.5456	1.7765	1.8830	0.7299
HOA-VMD-GRU	1.6854	0.7240	1.2982	0.8716
HOA-VMD-fixed ensemble	1.4458	0.9479	1.2024	0.8899
HOA-VMD-DWE	0.5087	0.5080	0.7133	0.9612

Forecasting under typical strong-gust conditions

Under typical strong-gust conditions, the rate of change in wind power output increases substantially, and the fluctuation characteristics become more pronounced. After processing the dataset corresponding to strong-gust conditions, the predicted values and error curves of different models are shown in *Figures 8* and *9*, respectively, while the comparative performance evaluation results are summarized in *Table 2*.

Based on the analysis of the results in *Figures 9* and *10*, the three single models (GRU, MLP, and LSTM) and the basic ensemble model exhibit considerable deviations between predicted values and actual measurements under typical strong gust conditions, demonstrating clearly insufficient fitting accuracy. The performance of the day-ahead 24-h persistence method is particularly limited in this scenario. Its prediction results fail to effectively track the drastic power fluctuations caused by instantaneous strong gusts, showing obvious trend mismatch and magnitude distortion. Consequently, prediction

errors are significantly amplified, making it difficult to meet the forecasting requirements of high-volatility scenarios. In contrast, the four models optimized by HOA-VMD demonstrate superior predictive capabilities across all evaluation metrics. Among them, the HOA-VMD-DWE model achieves the lowest prediction error, with its error curve showing the highest alignment with the zero-error line. This fully demonstrates the model's significant advantage in capturing the power fluctuation characteristics of typical strong gusts.

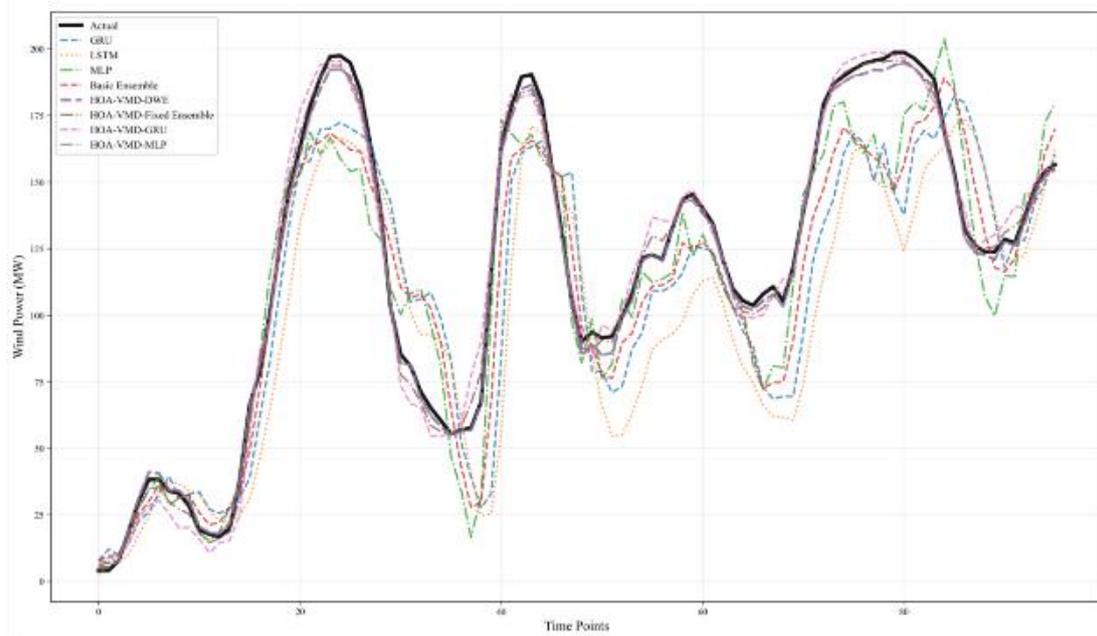


Figure 9. Prediction vs. actual values under strong-gust conditions

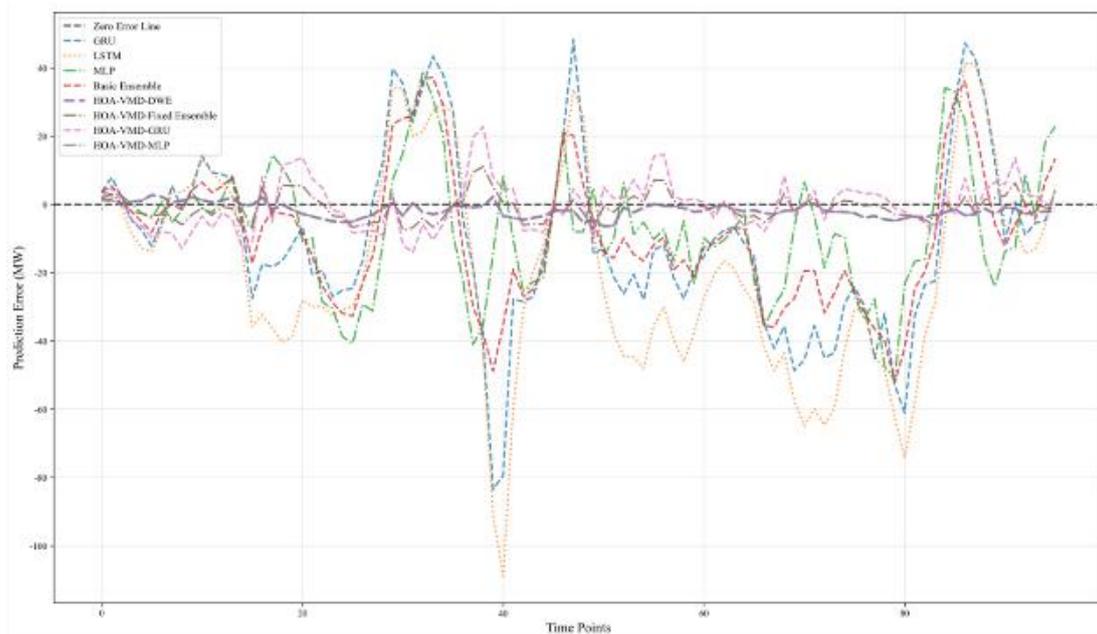


Figure 10. Prediction error curves under strong-gust conditions

Table 2. Error evaluation metrics of different models under strong-gust conditions

Model	MSE	MAE	RMSE	R ²
GRU	1261.7235	29.1583	35.5207	0.6079
LSTM	782.7786	22.6051	27.9782	0.7567
Basic ensemble	456.9080	17.4392	21.3754	0.8580
MLP	386.8256	15.4731	19.6679	0.8798
HOA-VMD-MLP	52.6028	5.8071	7.2528	0.9837
HOA-VMD-GRU	15.1042	3.0652	3.8864	0.9953
HOA-VMD- fixed ensemble	7.0395	2.2012	2.6532	0.9978
HOA-VMD-DWE	7.0390	2.1999	2.6531	0.9978

According to the analysis of *Figure 11*, the DWE mechanism also displays significant and more aggressive adaptive adjustment characteristics under high-volatility conditions. During the rapid onset and sustained duration of strong gusts, the weight of the MLP model remains almost consistently dominant and significantly above the equal-weight reference line. Meanwhile, the weight of the GRU model is compressed to a lower level, showing only a brief rise during the initial abrupt change phase of the gust. This weight distribution characteristic indicates that under strong gust weather, the wind power sequence exhibits significant high-frequency fluctuations and abrupt magnitude changes. Its variation patterns are dominated more by instantaneous nonlinear mapping relationships, while temporal dependencies and historical inertial features are relatively weakened. In this context, MLP demonstrates higher instantaneous fitting capability in capturing the strong nonlinear relationships between input and output, and is therefore continuously assigned higher weights by the DWE mechanism.

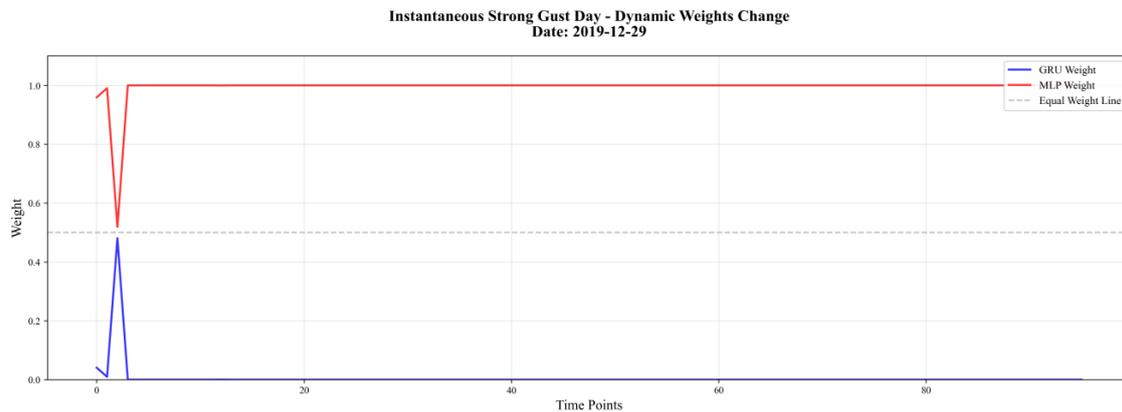


Figure 11. DWE weight curves under strong-gust conditions

It is worth noting that the short-term rise in GRU weight at specific moments usually corresponds to the initial stage of the gust or the transition process of the local fluctuation structure from disorder to phasic stability. This reflects that DWE does not completely ignore the contribution of the temporal model; instead, it dynamically introduces its memory advantage based on changes in the local structure of the wind power sequence. This rapid weight switching behavior under extreme fluctuation conditions fully

illustrates that DWE can perceive the source of prediction difficulties in real-time during non-stationary conditions such as strong gusts, and make targeted scheduling at the model level. Overall, these results further validate DWE's capability for precise identification of model advantages in high-disturbance and high-nonlinear scenarios, providing key support for enhancing the stability and reliability of wind power forecasting under extreme weather conditions.

According to *Table 2*, the optimized models—HOA-VMD-GRU, HOA-VMD-Fixed Ensemble, HOA-VMD-MLP, and HOA-VMD-DWE—consistently outperform the single models and the baseline ensemble in terms of MSE, MAE, and RMSE. In particular, the HOA-VMD-MLP and HOA-VMD-DWE models achieve the most significant improvements, with both attaining an R^2 of 0.9978, closest to 1. Furthermore, the HOA-VMD-DWE model records a lower MAE compared to HOA-VMD-MLP, indicating slightly better comprehensive forecasting performance. Overall, the HOA-VMD-DWE model exhibits outstanding accuracy in capturing wind power fluctuations under strong-gust conditions.

Normal wind weather prediction

Under normal wind conditions, the wind power output exhibits moderate fluctuations, distinct from the rapid variations observed during strong-gust conditions and the low-amplitude variations under weak-wind conditions. After processing the dataset corresponding to normal wind conditions, the predicted values and error curves of different models are shown in *Figures 10* and *11*, respectively, while the performance evaluation results are summarized in *Table 3*.

Based on the analysis of the results in *Figures 12* and *13*, the three single models (GRU, MLP, and LSTM) and the basic ensemble model still exhibit certain deviations between predicted values and actual measurements under general wind conditions, indicating insufficient overall fitting accuracy. The day-ahead 24-h persistence method demonstrates a certain degree of phasic applicability in this scenario. It maintains good fitting performance during intervals with small power fluctuation amplitudes and relatively gentle variations. However, during periods within general wind weather where wind speed levels are relatively higher and power fluctuations intensify, its prediction results exhibit significant response lag and error accumulation, making it difficult to accurately characterize the actual trend of power variations.

Table 3. Error evaluation metrics of different models under normal wind conditions

Model	MSE	MAE	RMSE	R^2
GRU	213.1897	10.0454	14.6010	0.2671
LSTM	139.1936	9.0643	11.7981	0.5215
Basic ensemble	83.7834	6.6965	9.1533	0.7120
MLP	70.2981	5.5923	8.3844	0.7583
HOA-VMD-MLP	7.4658	2.1314	2.7324	0.9743
HOA-VMD-GRU	3.7279	1.5561	1.9308	0.9872
HOA-VMD-fixed ensemble	3.5177	1.5594	1.8755	0.9879
HOA-VMD-DWE	2.7277	1.3796	1.6516	0.9906

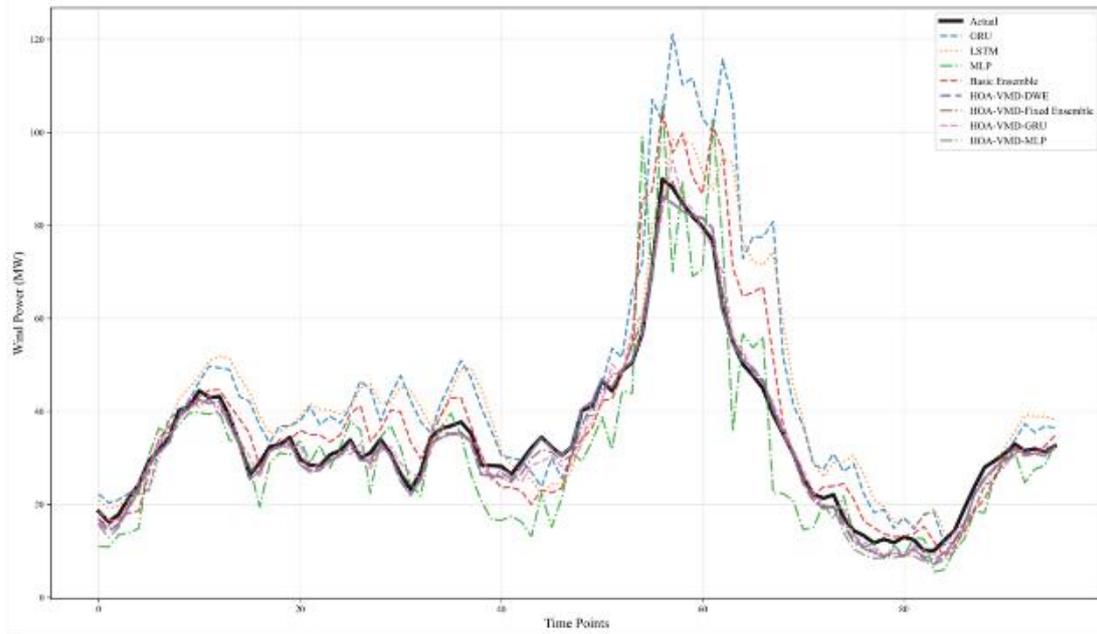


Figure 12. Prediction vs. actual values under normal wind conditions

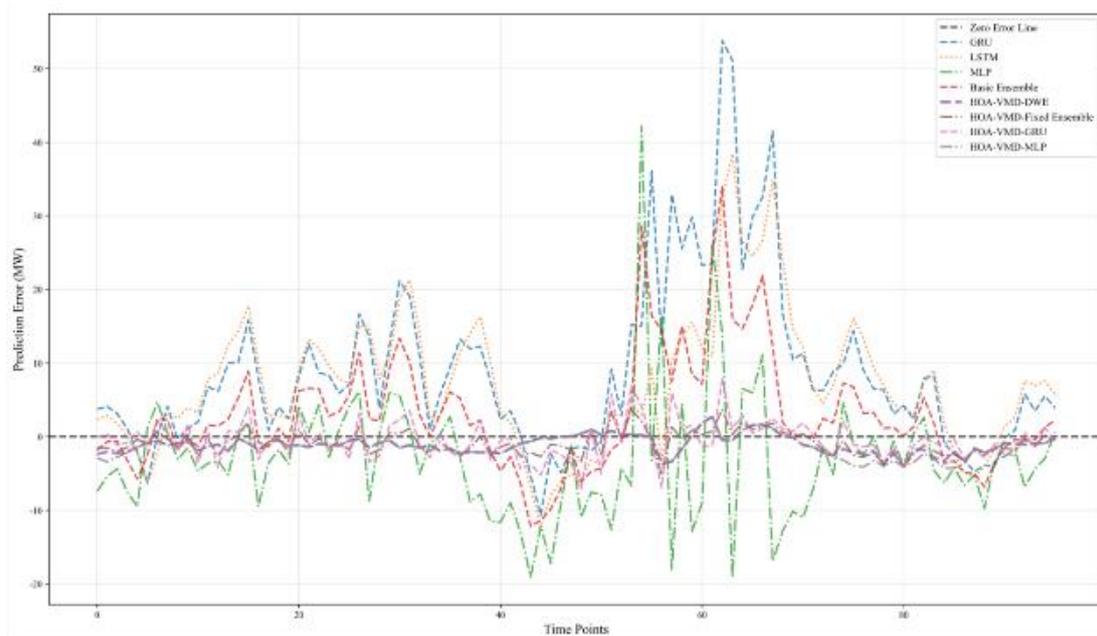


Figure 13. Prediction error curves under normal wind conditions

In contrast, the four models optimized by HOA-VMD demonstrate superior predictive performance across all evaluation metrics. Among them, the HOA-VMD-DWE model achieves the lowest prediction error, with its error curve showing the highest alignment with the zero-error line. This fully demonstrates the model's significant advantage in capturing wind power fluctuation characteristics under general wind conditions.

According to the analysis of *Figure 14*, the DWE mechanism exhibits a more balanced weight allocation characteristic with structure-aware capabilities under moderate wind

speed conditions. During phases where wind speed levels are relatively stable and power variation magnitude dominates, the weight of the MLP model is maintained at a high level for extended periods. This indicates its strong capability in characterizing the nonlinear mapping relationships of wind power under moderate wind conditions.

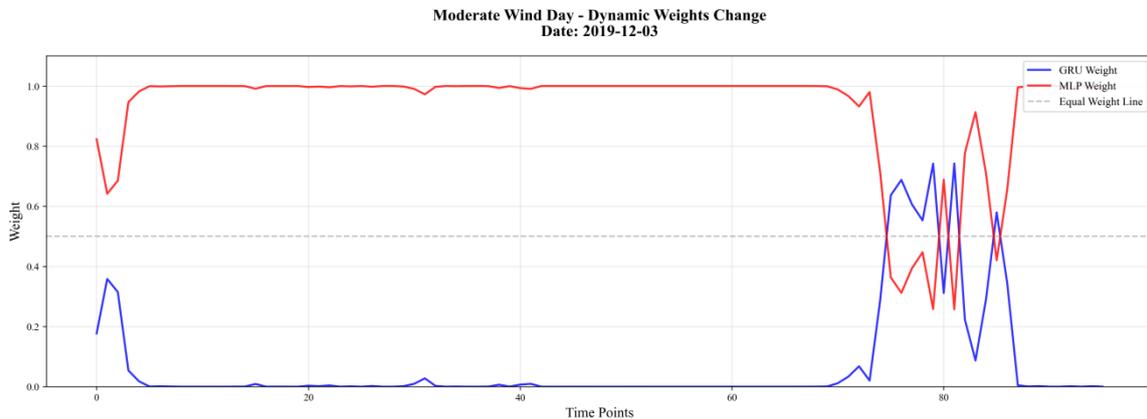


Figure 14. DWE weight curves under normal wind conditions

However, it is worth noting that during the relatively weak wind phases within general wind weather, the weight proportion of the GRU model shows a distinct increase. This phenomenon suggests that when wind speed decreases and power variations tend to become gentle, the temporal continuity and inertial features of the wind power sequence regain dominance. Consequently, GRU’s advantages in modeling historical dependencies and dynamic evolution patterns are fully realized, leading to an adaptive weight increase by the DWE mechanism. The aforementioned weight variation process reflects that DWE can keenly distinguish the dynamic characteristics of different sub-stages under a moderate wind background. It achieves refined scheduling between GRU and MLP, providing effective support for enhancing the overall stability and adaptability of wind power forecasting under general wind conditions.

According to *Table 3*, the optimized models—HOA-VMD-GRU, HOA-VMD-Fixed Ensemble, HOA-VMD-MLP, and HOA-VMD-DWE—significantly outperform the single models and the baseline ensemble across all error metrics. Notably, the HOA-VMD-DWE model demonstrates the best overall performance, achieving the lowest MSE, MAE, and RMSE values. Its coefficient of determination (R^2) reaches 0.9906, closest to 1, confirming its capability to accurately capture wind power output trends and achieve high-precision forecasting under normal wind conditions.

Conclusion

This paper addresses the challenges posed by the strong non-stationarity and complex fluctuations of wind power output sequences and proposes a short-term wind power forecasting method based on HOA-VMD-DWE. First, the Variational Mode Decomposition (VMD) algorithm is employed to decompose the original power sequence into intrinsic mode functions (IMFs) with different frequency-band characteristics. This significantly enhances the time–frequency resolution and hierarchical representation of features, thereby providing more refined inputs for subsequent modeling. Meanwhile, the

Hiking Optimization Algorithm (HOA) is introduced to perform global adaptive optimization of the key hyperparameters of VMD. This approach improves the globality and efficiency of the search process while avoiding the local optimum problem inherent in conventional optimization methods, thereby ensuring the stability of decomposition results and enhancing the robustness and generalization capability of the model under different operating conditions. Furthermore, the Dynamic Weight Ensemble (DWE) algorithm is utilized to dynamically integrate the prediction results of the Multilayer Perceptron (MLP) and the Gated Recurrent Unit (GRU) models. By adaptively adjusting the contributions of the two models according to sequence features, the method not only leverages the GRU's advantage in modeling temporal dependencies but also exploits the MLP's strength in nonlinear relationship fitting, thereby achieving more accurate fitting and prediction of wind power sequences. Experimental results demonstrate that the HOA-VMD-DWE model achieves significantly better forecasting performance under various representative operating conditions compared with conventional single models and basic ensemble models, fully highlighting its advantages in handling complex nonlinear relationships and temporal dependencies.

In conclusion, the wind power forecasting method based on HOA-VMD-DWE exhibits excellent predictive performance and high practical value. It can provide more scientific decision support for power sector operations, thereby optimizing scheduling and resource allocation, reducing operational costs, and improving the overall efficiency of power system operation.

REFERENCES

- [1] Ahmadi, M., Aly, H., Khashei, M. (2025): Enhancing power grid stability with a hybrid framework for wind power forecasting: integrating Kalman filtering, deep residual learning, and bidirectional LSTM. – *Energy* 137752. <https://doi.org/10.1016/j.energy.2025.137752>.
- [2] Duan, X. (2025): Settlement prediction of Nanjing Metro Line 10 with HOA-VMD-LSTM. – *Measurement* 244: 116477. <https://doi.org/10.1016/j.measurement.2024.116477>.
- [3] Hou, G., Wang, J., Fan, Y. (2024): Multistep short-term wind power forecasting model based on secondary decomposition, the kernel principal component analysis, an enhanced arithmetic optimization algorithm, and error correction. – *Energy* 286: 129640. <https://doi.org/10.1016/j.energy.2023.129640>.
- [4] Huang, J., Zhang, W., Qin, J., Song, S. (2024): Ultra-short-term wind power prediction based on eEEMD-LSTM. – *Energies* 17(1):251. <https://doi.org/10.3390/en17010251>.
- [5] Khan, S., Muhammad, Y., Jadoon, I., Awan, S. E., Raja, M. A. Z. (2025): Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting. – *Applied Soft Computing* 170: 112765. <https://doi.org/10.1016/j.asoc.2025.112765>.
- [6] Li, L., Han, S., Wang, Y. M. (2013): A physical approach of the short-term wind power prediction based on CFD pre-calculated flow fields. – *Journal of Hydrodynamics, Ser. B* 25(1): 56-61. [https://doi.org/10.1016/S1001-6058\(13\)60338-8](https://doi.org/10.1016/S1001-6058(13)60338-8).
- [7] Li, Y., Sun, K., Yao, Q., Wang, L. (2024): A dual-optimization wind speed forecasting model based on deep learning and improved dung beetle optimization algorithm. – *Energy* 286: 129604. <https://doi.org/10.1016/j.energy.2023.129604>.
- [8] Lin, G., Chi, Y., Ding, X., Zhang, Y., Wang, J., Wang, C., et al. (2025): Wind-speed prediction in renewable-energy generation using an IHOA. – *Sustainability* 17(14): 6279. <https://doi.org/10.3390/su17146279>.

- [9] Liu, M., Wang, L., Pang, X., Zheng, Z., Li, H. (2025): Hierarchical model based on optimized feature decomposition and deep learning for short-term wind-power forecasting. – *International Journal of Electrical Power & Energy Systems* 172: 111097. <https://doi.org/10.1016/j.ijepes.2025.111097>.
- [10] Liu, Y., Guan, L., Hou, C., Han, H., Liu, Z., Sun, Y., Zheng, M. (2019): Wind power short-term prediction based on LSTM and discrete wavelet transform. – *Applied Sciences* 9(6): 1108. <https://doi.org/10.3390/app9061108>.
- [11] Qin, G., Yan, Q., Zhu, J., Xu, C., Kammen, D. M. (2021): Day-ahead wind power forecasting based on wind load data using hybrid optimization algorithm. – *Sustainability* 13(3): 1164. <https://doi.org/10.3390/su13031164>.
- [12] Rayi, V. K., Mishra, S. P., Naik, J., Dash, P. K. (2022): Adaptive VMD based optimized deep learning mixed kernel ELM autoencoder for single and multistep wind power forecasting. – *Energy* 244: 122585. <https://doi.org/10.1016/j.energy.2021.122585>.
- [13] Ribeiro, M. H. D. M., Da Silva, R. G., Moreno, S. R., Canton, C., Larcher, J. H. K., Stefenon, S. F., et al. (2024): Variational mode decomposition and bagging extreme learning machine with multi-objective optimization for wind power forecasting. – *Applied Intelligence* 54(4): 3119-3134. <https://doi.org/10.1007/s10489-024-05331-2>.
- [14] Sireesha, P. V., Thotakura, S. (2024): Wind power prediction using optimized MLP-NN machine learning forecasting model. – *Electrical Engineering* 106(6): 7643-7666. <https://doi.org/10.1007/s00202-024-02440-6>.
- [15] Stathopoulos, C., Kaperoni, A., Galanis, G., Kallos, G. (2013): Wind power prediction based on numerical and statistical models. – *Journal of Wind Engineering and Industrial Aerodynamics* 112: 25-38. <https://doi.org/10.1016/j.jweia.2012.09.004>.
- [16] Wang, Y., Song, Z., Liu, Y., Chen, J. (2022): Wind power Prediction based on the fusion of CN N-GRU combined neural network and attention mechanism. – In: 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, pp. 698-702. <https://doi.org/10.1109/IAEAC54830.2022.9929869>.
- [17] Xin, Z., Chi, W., Zhang, H., Liu, X., Zheng, G., Sun, N., et al. (2025): Ultra short-term wind power prediction based on an improved temporal convolutional network. – *International Journal of Green Energy* 1-21. <https://doi.org/10.1080/15435075.2025.2477627>.
- [18] Xu, X., Cao, Q., Deng, R., Guo, Z., Chen, Y., Yan, J. (2025): A cross-dataset benchmark for neural network-based wind power forecasting. – *Renewable Energy* 123463. <https://doi.org/10.1016/j.renene.2025.123463>.
- [19] You, G. D., Chang, Z. C., Li, X. Y., Liu, Z. F., Xiao, Z. Y., Lu, Y. R., Zhao, S. (2024): Using enhanced variational modal decomposition and dung beetle optimization algorithm optimization-kernel extreme learning machine model to forecast short-term wind power. – *Electric Power Systems Research* 236: 110904. <https://doi.org/10.1016/j.epsr.2024.110904>.
- [20] Yu, H., Dai, Q. (2022): DWE-IL: a new incremental learning algorithm for non-stationary time series prediction via dynamically weighting ensemble learning. – *Applied Intelligence* 52(1): 174-194. <https://doi.org/10.1007/s10489-021-02385-4>.
- [21] Zhang, Y., Zhang, L., Sun, D., Jin, K., Gu, Y. (2023): Short-term wind power forecasting based on VMD and a hybrid SSA-TCN-BiGRU network. – *Applied Sciences* 13(17): 9888. <https://doi.org/10.3390/app13179888>.
- [22] Zhang, Z., Wei, Z., Nie, B., Li, Y. (2022): Discontinuous maneuver trajectory prediction based on HOA-GRU method for the UAVs. – *Electronic Research Archive* 30(8). <https://doi.org/10.3934/era.2022158>.
- [23] Zhou, H., Jiang, J. X., Huang, M. (2011): Short-term wind power prediction based on statistical clustering. – In: 2011 IEEE Power and Energy Society General Meeting. IEEE, pp. 1-7. <https://doi.org/10.1109/PES.2011.6039233>.

- [24] Zhou, Z., Li, X., Wu, H. (2016): Wind power prediction based on random forests. – In: 2016 4th International Conference on Electrical & Electronics Engineering and Computer Science (ICEEECS 2016). Atlantis Press, Paris, pp. 352-356.
<https://doi.org/10.2991/iceeeecs-16.2016.73>.