



ARTICLE

Ultra-Short-Term Wind Power Forecasting Based on Hierarchical Signal Refinement and Intelligently Optimized Deep Learning

Xiaolan Li^{1,2,*}, Jinyu Shen^{1,2}, Jinhuang Liang^{1,2} and Yanting Wang^{1,2}

¹School of Electrical and Information Technology, Yunnan Minzu University, Kunming, China

²Yunnan Key Laboratory of Unmanned Autonomous System, Kunming, China

*Corresponding Author: Xiaolan Li. Email: lixiaolan@ymu.edu.cn

Received: 22 November 2025; Accepted: 05 January 2026; Published: 18 June 2026

ABSTRACT: The intrinsic volatility and stochasticity of large-scale wind power generation pose significant challenges to grid stability. To address the limitations of conventional models in capturing strong non-stationarity, this study proposes a novel Multi-Stage Adaptive Forecasting Network (MSAF-Net). The framework features a hierarchical signal refinement strategy coupled with an intelligently optimized hybrid predictor. Initially, input redundancy is minimized via Pearson Correlation Coefficient (PCC) analysis to isolate significant meteorological variables. A two-phase decomposition-reconstruction mechanism is then implemented: the wind power series is first decomposed using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). To optimize the trade-off between signal complexity and computational cost, the resulting components are reconstructed based on Sample Entropy (SE), with the highest-complexity component specifically targeted for secondary denoising via Empirical Wavelet Transform (EWT). For the prediction stage, a hybrid architecture integrates Bidirectional Temporal Convolutional Networks (BiTCN) to extract multi-scale local features and Bidirectional Long Short-Term Memory (BiLSTM) networks to model long-term temporal dependencies. Crucially, an Attention Mechanism is embedded to weigh critical time steps, while the Sparrow Search Algorithm (SSA) automatically optimizes the network hyperparameters. Experimental results demonstrate that MSAF-Net achieves an RMSE of 41.59, MAE of 26.67, and MAPE of 1.36%. Notably, the proposed model achieves a 23.16% reduction in MAPE compared to the competitive CEEMDAN-EWT-LSTM benchmark, verifying its superior predictive accuracy and generalization capability.

KEYWORDS: Two-layer decomposition; hybrid model; SSA-BiLSTM forecasting; restructure

1 Introduction

The escalating global demand for energy, coupled with the imperative for environmental sustainability, has accelerated the transition towards renewable energy sources. Wind energy, characterized by its cleanliness and cost-competitiveness, has emerged as a pivotal component of the global energy portfolio [1]. Furthermore, economic projections indicate a substantial reduction in the levelized cost of wind power by 2050, further solidifying its role in future power systems [2]. However, the intrinsic intermittency and stochastic volatility of wind power present significant challenges [3]. The reliance on fluctuating meteorological conditions leads to erratic power output, introducing technical complexities and threatening the stability of grid operations [4].

To improve the efficiency and reliability of wind power generation, numerous studies have investigated advanced control strategies and modeling approaches for wind turbines [5,6]. Representative techniques



include Gaussian Process Regression (GPR), Autoregressive Moving Average (ARMA), and Kalman Filter (KF) methods. For instance, Jin et al. [7] utilized an adaptive combination of Finite Mixture Gaussian Process Regression models to generate wind power forecasts. Kong et al. [5] introduced a Distributed Economic Model Predictive Control strategy that incorporates both power output tracking and economic optimization for wind farm management. Aly [8] constructed a three-stage wind prediction framework integrating an Adaptive Neuro-Fuzzy Inference System, a Recurrent Kalman Filter, and a Wavelet Neural Network, thereby identifying an optimal hybrid model. Nevertheless, approaches that depend exclusively on historical data often struggle to capture the intricate nonlinear patterns inherent in wind power sequences, and predictive performance tends to deteriorate as the forecast horizon extends [9].

In contrast to conventional statistical methods that depend solely on historical data, artificial intelligence-based models leverage the mapping relationships between Numerical Weather Prediction (NWP) data and historical wind power generation to forecast future output. These models demonstrate a strong capacity for nonlinear fitting, enabling more accurate and adaptive predictions under complex meteorological conditions [10]. Commonly employed AI techniques include the Long Short-Term Memory (LSTM) network and the Temporal Convolutional Network (TCN). The distinctive gated structure of LSTM enables the propagation of information from earlier time steps to later ones within a sequence [11]. Mohsen et al. [12] utilized LSTM to process complex non-stationary time-series signals, validating its effectiveness in mitigating the vanishing gradient problem. Nevertheless, conventional unidirectional LSTM and TCN architectures capture only forward temporal dependencies, overlooking backward influences in time series [13]. To address this limitation, Graves and Schmidhuber [14] introduced the Bidirectional LSTM (BiLSTM), which explicitly models both forward and backward temporal relationships. Experimental results have confirmed that BiLSTM achieves higher predictive accuracy than the standard LSTM architecture. Nevertheless, a significant gap persists in the existing literature regarding the optimal configuration of BiLSTM network parameters. These parameters, which are critical components of neural network architectures, govern essential characteristics such as learning rate and model capacity. Suboptimal parameter settings may hinder the network's ability to adequately fit the training data, thereby limiting its predictive performance. To address this challenge, intelligent optimization algorithms are increasingly adopted as a dedicated optimization layer to identify ideal parameter configurations for both LSTM and BiLSTM networks. For instance, Fan et al. [15] employed the Grey Wolf Optimizer (GWO) to optimize BiLSTM hyperparameters for power load forecasting. In [16], the PSO algorithm was implemented to tune critical parameters of BiLSTM, reporting a 2% increase in predictive accuracy over alternative methods; however, this study did not incorporate the influence of meteorological variables on PV forecasting. The authors of [17] utilized the Whale Optimization Algorithm (WOA) to determine optimal values for the initial learning rate and the maximum number of iterations in a BiLSTM network. A more recent contribution is the Sparrow Search Algorithm (SSA), introduced in 2020 [18]. Comparative simulations revealed that SSA exceeds contemporary algorithms in convergence speed, solution precision, stability, and robustness. It is especially effective in multi-objective optimization problems, demonstrating rapid convergence and high precision, which makes it particularly well-suited for optimizing hyperparameters in BiLSTM networks.

Moreover, the predictive accuracy of individual models is intrinsically constrained. Limouni et al. [19] developed a hybrid photovoltaic forecasting model incorporating a TCN-LSTM network that integrates meteorological factors. Their results indicated that the combined TCN-LSTM structure outperformed both standalone TCN and LSTM models. While the aforementioned TCN-LSTM architecture was designed for photovoltaic scenarios, its ability to extract temporal dependencies is theoretically transferable to wind power tasks. Given the greater chaotic characteristics of wind speed compared to solar radiation, this study adapts and enhances this hybrid architecture specifically for wind power forecasting. To verify this potential

in wind power prediction, Chen et al. [20] proposed a VMD-guided hybrid framework (VMD-GDPSO-TCN-BiLSTM). This model addresses the non-stationarity of wind speed series by integrating TCN's causal convolution with BiLSTM's bidirectional modeling. Their results showed superior generalization and higher accuracy compared to standard TCN-LSTM benchmarks.

Inspired by this approach, the proposed BiTCN-SSA-BiLSTM-Attention model targets two primary objectives. First, it captures bidirectional temporal dependencies by capitalizing on the enhanced capability of hybrid architectures. Second, it employs SSA for the effective, automated optimization of BiLSTM hyperparameters. Furthermore, the incorporated attention mechanism dynamically emphasizes the most relevant segments of the input sequence, filtering out irrelevant noise and distractions, thereby enabling precise identification of decisive “critical moments” that significantly influence forecasting results [21].

Owing to the intermittent characteristics of wind, wind power generation data are highly stochastic and volatile [22]. These attributes pose considerable challenges to achieving accurate predictions using a single forecasting methodology. In response, researchers have incorporated data decomposition techniques as a preprocessing step, which has yielded encouraging outcomes [23]. In such decomposition-based hybrid forecasting frameworks, the original wind power signal is first broken down into multiple relatively stable subseries. Individual prediction models are subsequently constructed for each subseries, and their outputs are aggregated to produce the final forecast. For instance, Yang et al. [24] applied Empirical Mode Decomposition (EMD) to decompose the raw data, utilized BO to tune LSTM hyperparameters, and predicted each intrinsic mode function (IMF) using LSTM before reconstructing the final output through summation. Chen et al. [25] adopted Ensemble Empirical Mode Decomposition (EEMD) combined with a Genetic Algorithm-optimized LSTM model, reporting high predictive accuracy. A notable limitation of EEMD, however, is its dependence on empirical selection of white noise amplitude, which can lead to mode mixing and compromise decomposition fidelity. To mitigate this issue, Fang et al. [26] introduced a three-layer forecasting architecture incorporating outlier detection, Empirical Wavelet Transform (EWT), and an ensemble of neural networks. Their results confirmed EWT's efficacy in decomposition and a substantial improvement in prediction accuracy. Among these techniques, CEEMDAN has been shown to outperform EMD, EEMD, and CEEMD, offering superior decomposition performance [27]. However, the highest-frequency component—the first IMF produced by CEEMDAN remains highly volatile and contaminated with noise [28]. This component poses the greatest prediction challenge [29] and can substantially compromise overall forecasting accuracy. To address this, Karijadi et al. [30] decomposed wind power data using CEEMDAN, denoised the first IMF via EWT, predicted each IMF using LSTM, and aggregated the results, effectively mitigating the impact of noise. Li et al. [31] employed CEEMDAN for decomposition and utilized Sample Entropy (SE) to classify and reconstruct IMFs, thereby reducing model complexity and streamlining the prediction process. Similarly, Su et al. [32] constructed a hybrid system using CEEMDAN and VMD-based secondary denoising, enhancing the generalization of subsequent Transformer-GRU predictors. Zhou et al. [33] adopted CEEMDAN for signal decomposition, applied TCN to extract spatial correlations between meteorological variables and wind speed, and incorporated an Attention Mechanism (AM) with BiLSTM to capture temporal dependencies. Their experimental results demonstrated that this model achieved high predictive accuracy across varying spatial scales. Despite these advances, substantial noise remains present in the low-order modes derived from CEEMDAN, which continues to significantly impair prediction performance.

Despite these advances, substantial noise remains present in the low-order modes derived from CEEMDAN, which continues to significantly impair prediction performance.

While recent advancements have enhanced forecasting capabilities, fundamental structural limitations persist, restricting model reliability in dynamic real-world scenarios. As critically analyzed in Table 1, these

deficiencies primarily manifest in three aspects. Specifically, regarding decomposition, most hybrid models rely on single-stage methods (e.g., only CEEMDAN), where the highest-frequency component (IMF1) retains significant stochastic volatility [28]. Without targeted secondary refinement, this residual noise propagates through the predictor, inevitably compromising overall accuracy. Furthermore, conventional architectures often lack a holistic temporal view; standard TCNs fail to capture backward dependencies [34], while basic BiLSTMs struggle to distinguish critical “turning points” from normal fluctuations without an attention mechanism, leading to information overload and tracking lag. Finally, concerning optimization, many complex frameworks (e.g., [32]) still depend on manual hyperparameter tuning or basic algorithms. This static configuration creates a high risk of falling into local optima and limits the model’s ability to generalize across varying wind regimes.

Table 1: Comparison of the proposed framework with representative existing methods, highlighting specific methodological gaps

Study	Core architecture	Decomposition	Denoising	Optimization	Critical limitation & consequence
Jin et al. [7]	GPR mixture	None	No	Adaptive	Limited nonlinear mapping capability; poor performance in highly volatile wind regimes.
Aly [8]	ANFIS-RKF	Wavelet	Yes	None	Wavelet basis limits adaptivity; lack of parameter optimization leads to suboptimal fitting.
Yang et al. [24]	LSTM	EMD	No	BO	Single EMD suffers from mode mixing; residual noise in IMF1 degrades prediction stability.
Zhou et al. [33]	TCN-BiLSTM-AM	CEEMDAN	No	RIME	Failure to separate high-frequency stochastic noise; prediction lag occurs during rapid power ramps.
Karijadi et al. [30]	LSTM	CEEMDAN+EWT	Yes	None	Unidirectional LSTM misses backward dependencies; absence of adaptive optimization limits generalization.
Su et al. [32]	Transformer-GRU	CEEMDAN+VMD	Yes	Manual	Reliance on manual hyperparameter setting; high risk of falling into local optima or overfitting.
MSAF-Net	BiTCN-BiLSTM-Attn	CEEMDAN+EWT	Targeted	SSA	Proposed Solution: Hierarchical refinement eliminates noise; SSA ensures global optimality; Dual-context learning captures trends.

To systematically overcome these inherent limitations, this study proposes a novel hybrid framework termed the Multi-Stage Adaptive Forecasting Network (MSAF-Net). The architecture is founded on a divide-and-conquer philosophy, integrating a hierarchical signal refinement pipeline with an intelligently optimized deep learning predictor.

In the preprocessing phase, the framework balances signal stationarity and computational efficiency. Instead of applying blanket decomposition, we implement a complexity-aware strategy. First, the non-stationary wind power series is stabilized via CEEMDAN. Next, SE serves as a quantitative metric to assess IMF complexity. Based on SE values, similar modes are merged to reduce dimensionality. Finally, EWT is selectively applied only to the highest-complexity component for secondary denoising. This targeted refinement suppresses high-frequency stochastic volatility without the computational overhead of full-spectrum secondary decomposition.

For the prediction phase, the framework constructs an adaptive engine. It leverages a BiTCN to capture multi-scale local dependencies from both forward and retrospective contexts. These features are then integrated with a BiLSTM to model global temporal trends. To further enhance robustness, a dual-adaptivity mechanism is embedded: an Attention Module dynamically highlights salient time steps, while the SSA autonomously navigates the hyperparameter space. The principal contributions of this study are condensed as follows:

- (1) A hierarchical signal refinement strategy is proposed. By integrating CEEMDAN with SE-based reconstruction, the model aggregates similar modes to reduce dimensionality. Additionally, EWT is selectively applied to the highest-complexity component, providing targeted denoising without over-processing stable signals.
- (2) A bidirectional convolutional architecture improves contextual feature extraction. The BiTCN module expands the receptive field to capture information flows in both directions, improving the identification of deep correlations between wind power and meteorological variables.
- (3) An adaptive hybrid predictor featuring swarm intelligence is constructed for robust forecasting. The framework synergizes BiLSTM with an Attention Mechanism to focus on critical temporal features. Moreover, the SSA is employed to autonomously optimize hyperparameters, eliminating manual tuning errors and ensuring high generalization capability under varying wind conditions.

2 Methodology

2.1 Hierarchical Signal Decomposition

To effectively manage the inherent non-stationarity and stochastic volatility of wind power signals, this study constructs a hierarchical signal processing mechanism. The framework initiates with Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for global trend extraction, followed by Empirical Wavelet Transform (EWT) for targeted secondary refinement.

CEEMDAN is first employed to decompose the non-linear wind power signal into a set of IMFs. As an advancement of the original EMD framework [35], CEEMDAN mitigates the critical issue of mode mixing by iteratively adding adaptive noise at each decomposition stage. This innovation ensures precise signal reconstruction and eliminates the residual noise common to EEMD, providing a more accurate and computationally efficient decomposition [36]. The detailed algorithm proceeds as follows:

First, add k times zero-mean white noise $w^i(t)$ to the original wind power signal $x(t)$, generating k new sequences $x^i(t)$:

$$x^i(t) = x(t) + \epsilon_0 w^i(t), \quad i = 1, 2, 3, \dots, L \quad (1)$$

EMD is then employed across all signal subsets to identify k inherent mode components, yielding the initial mode $IMF_1(t)$ and its residue $r_1(t)$:

$$IMF_1(t) = \frac{1}{k} \sum_{i=1}^k IMF_1^i(t) \quad (2)$$

$$r_1(t) = x(t) - IMF_1(t) \quad (3)$$

Subsequently, adaptive noise is added to the residuals recursively. For the g -th component, the intrinsic mode function $IMF_g(t)$ and residual $r_g(t)$ are extracted as:

$$IMF_g(t) = \frac{1}{k} \sum_{i=1}^k E_1(r_{g-1}(t) + \epsilon_{g-1} E_{g-1}(w^i(t))) \quad (4)$$

$$r_g(t) = r_{g-1}(t) - IMF_g(t) \quad (5)$$

This process repeats until the residual $R(t)$ reaches monotonicity, concluding the primary decomposition:

$$x(t) = \sum_{i=1}^k IMF_k(t) + R(t) \quad (6)$$

Although CEEMDAN effectively separates oscillation modes, high-frequency stochastic noise often persists in the initial IMFs. To mitigate this without the drawbacks of recursive filtering, EWT is employed for secondary signal refinement. Unlike EMD-based methods, EWT constructs an adaptive wavelet filter bank tailored to the specific frequency characteristics of the processed signal based on a solid mathematical foundation [37].

The core principle of EWT relies on the adaptive segmentation of the Fourier spectrum. The boundaries Ω_n of each segment are determined as the midpoints between two consecutive local maxima in the signal's spectrum:

$$\Omega_n = \frac{\omega_n + \omega_{n+1}}{2} \quad (7)$$

where ω_n denotes the frequency of the n -th local maximum. Based on these boundaries, a set of band-pass filters is implicitly defined using the Meyer wavelet construction to decompose the signal into sub-modes. The approximation trend and detail fluctuations are then extracted via the inverse Fourier transform:

$$\mathcal{W}_x^e(0, t) = \mathcal{F}^{-1} [\hat{x}(\omega) \cdot \hat{\phi}_1(\omega)] \quad (8)$$

$$\mathcal{W}_x^e(n, t) = \mathcal{F}^{-1} [\hat{x}(\omega) \cdot \hat{\psi}_n(\omega)], \quad n = 1, 2, \dots, N \quad (9)$$

Eqs. (8) and (9) formalize the reconstruction of time-domain components via the Inverse Fourier Transform (\mathcal{F}^{-1}). Specifically, Eq. (8) extracts the approximation coefficients $\mathcal{W}_x^e(0, t)$ by filtering the signal spectrum $\hat{x}(\omega)$ with the empirical scaling function $\hat{\phi}_1(\omega)$, thereby capturing the low-frequency deterministic trend. Conversely, Eq. (9) derives the detail coefficients $\mathcal{W}_x^e(n, t)$ using the adaptive wavelet functions $\hat{\psi}_n(\omega)$, which effectively isolate specific high-frequency bands containing stochastic volatility and transient features.

This secondary decomposition effectively isolates and removes stochastic noise from the deterministic signal features established in the first stage.

2.2 Bidirectional Temporal Convolutional Network (BiTCN)

TCN utilizes dilated causal convolutional layers that maintain consistent input and output lengths, effectively integrating the strengths of both convolutional and recurrent neural networks [38]. The inherent unidirectionality of standard TCNs is a critical flaw, as it precludes the capture of backward temporal information. To remedy this, we employ a BiTCN. This architecture processes the data stream from both directions, constructing a more holistic model of the long-range dependencies within the wind power time series. The architecture is illustrated in Fig. 1.

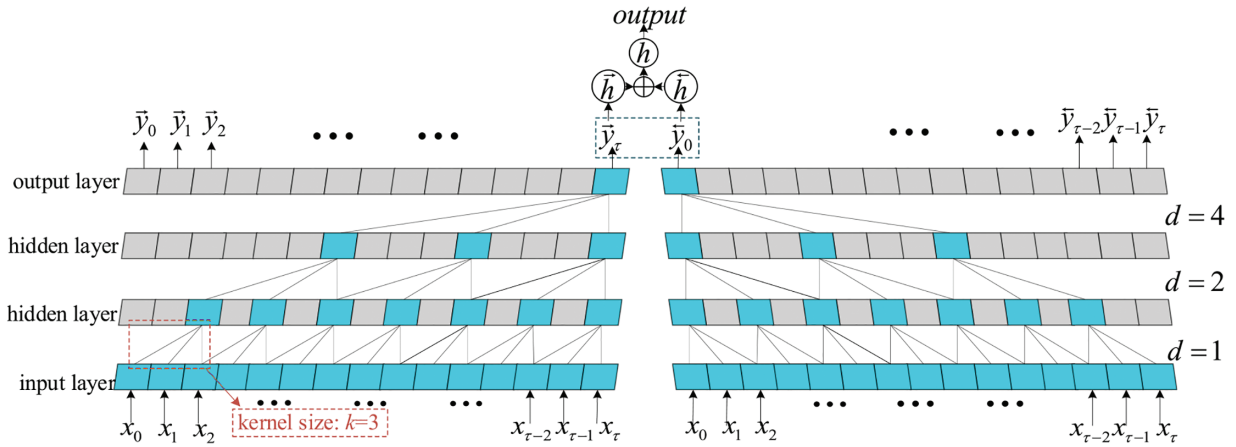


Figure 1: The structure of bidirectional dilated causal convolutional network

The BiTCN model incorporates dilated convolution, which enables exponential expansion of the receptive field with limited layers while preserving feature map dimensions. Given a one-dimensional input sequence $x \in R^n$ and a convolutional kernel filter: $\{0, \dots, k-1\} \rightarrow R$, the dilated convolution operation at element s is formally defined as:

$$F(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (10)$$

where k denotes the kernel size, d represents the dilation factor that controls the spacing between the points of the kernel (by inserting zeros), and the term $s - d \cdot i$ ensures the temporal dependence oriented to the past. The dilation factor d increases exponentially after each convolutional layer, allowing BiTCN to efficiently capture extended contextual information. However, enlarging the receptive field may lead to issues such as gradient vanishing and slow convergence. To mitigate these challenges while improving feature extraction, a residual block structure is integrated into the network, as illustrated in Fig. 2.

2.3 SSA-Optimized Bidirectional Long Short-Term Memory Network (SSA-BiLSTM)

While BiTCN excels at extracting local multi-scale features, it relies on subsequent layers to capture global long-term temporal dependencies. To achieve this, the feature sequence extracted by BiTCN is fed into a BiLSTM network.

Standard LSTM networks process data strictly in a forward chronological order, capturing only past information. However, wind power generation is a continuous physical process where the state at any given moment is correlated with the temporal context of the entire observation window. Although future data is unavailable during real-time inference, the bidirectional architecture allows the model to

learn representations from both past-to-future and future-to-past contexts during the training phase. This dual-context mechanism effectively reinforces the coherence of signal features and enables the model to better identify continuous trends amidst stochastic fluctuations, offering superior stability compared to unidirectional models.

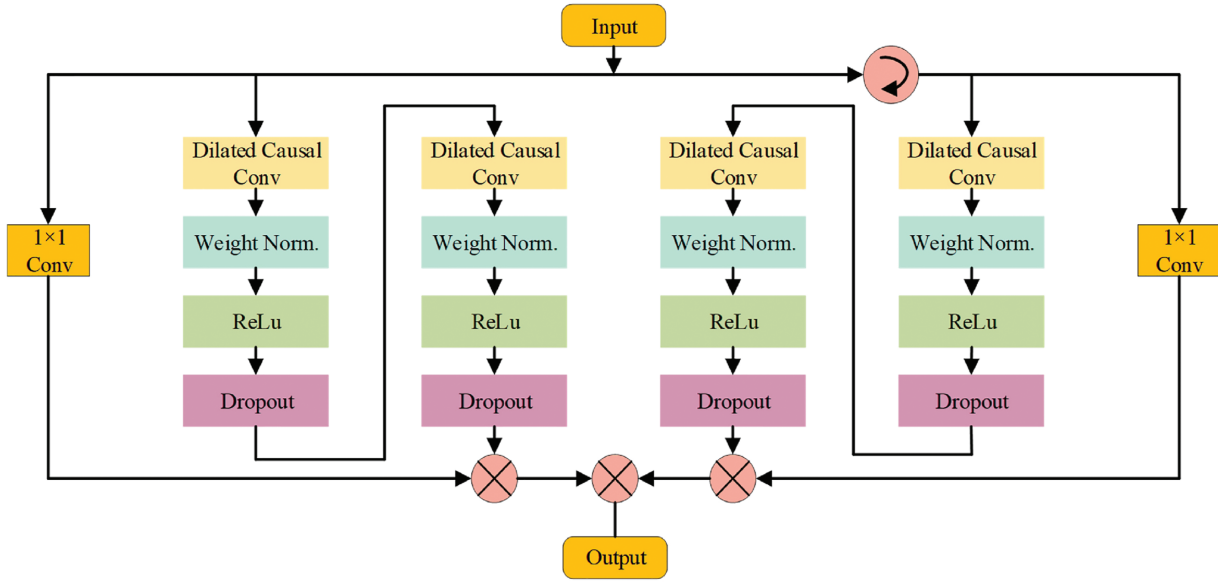


Figure 2: The structure of residual block in BiTCN

As illustrated in Fig. 3a, the LSTM unit utilizes gating mechanisms—specifically the forget, input, and output gates—to regulate information flow, thereby effectively mitigating the vanishing gradient problem inherent in standard RNNs. To capture the full temporal context, the BiLSTM architecture processes the input sequence in two opposite directions simultaneously, as shown in Fig. 3b. The final output state y_t is obtained by linearly combining the concatenated hidden states of the forward layer (\vec{h}_t) and the backward layer (\overleftarrow{h}_t):

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (11)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t+1}) \quad (12)$$

$$y_t = W_y[\vec{h}_t, \overleftarrow{h}_t] + b_y \quad (13)$$

where W_y and b_y represent the weight matrix and bias vector of the output layer, respectively. However, the predictive performance of the BiLSTM network is highly sensitive to its hyperparameter configuration, particularly the initial learning rate and the number of hidden units. Manual tuning of these parameters is often inefficient and prone to suboptimal local minima. To address this limitation, the SSA is integrated to establish an adaptive optimization framework.

In this proposed hybrid architecture, the mathematical optimization process is directly coupled with the neural network training. Specifically, each individual sparrow's position vector θ in the search space is mapped to a candidate set of BiLSTM hyperparameters. The quality of these parameters is evaluated via a fitness function, defined as the Root Mean Square Error (RMSE) of the model's prediction on the validation set:

$$F(\boldsymbol{\theta}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i(\boldsymbol{\theta}))^2} \quad (14)$$

where N denotes the number of validation samples and $\hat{y}_i(\boldsymbol{\theta})$ represents the prediction generated by the model configured with parameters $\boldsymbol{\theta}$. Throughout the iterative process, the SSA population—divided into producers, scroungers, and scouts—dynamically updates their positions to minimize this objective function $F(\boldsymbol{\theta})$. Upon convergence, the global optimal position $\boldsymbol{\theta}_{best}$ is decoded and applied to the final BiLSTM network, thereby automating the model tuning process and ensuring superior generalization capability.

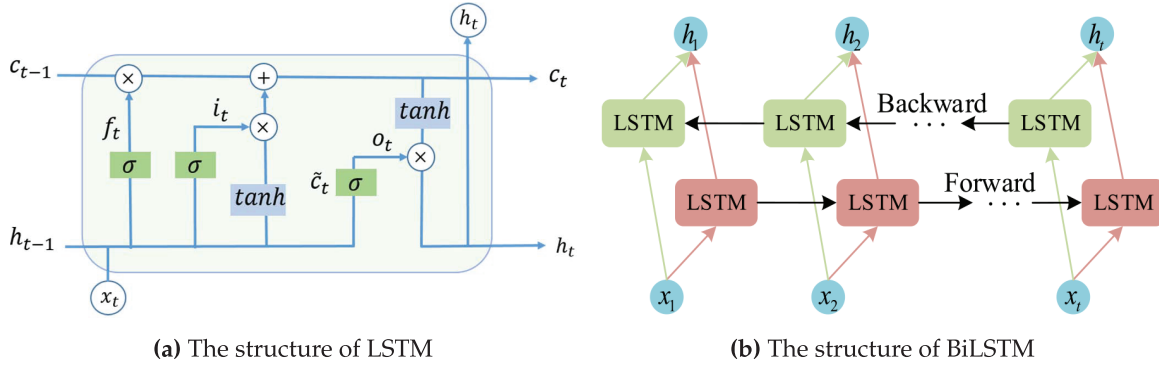


Figure 3: The bidirectional structure composed of LSTM units

3 Multi-Stage Adaptive Forecasting Network (MSAF-Net)

To mitigate the intrinsic non-stationarity and stochastic volatility of wind power, this paper proposes the Multi-Stage Adaptive Forecasting Network (MSAF-Net). The framework synergizes a dual-stage decomposition-reconstruction mechanism with an intelligently optimized deep learning predictor. Underpinned by a Coarse-to-Fine hierarchical refinement strategy, the architecture optimizes the critical trade-off between signal fidelity and computational efficiency. The process operates through a progressive filtering mechanism: CEEMDAN first decomposes the non-stationary series into intrinsic modes, after which SE serves as a quantitative metric to isolate the component exhibiting the highest complexity. Consequently, EWT is applied exclusively to this high-volatility component for secondary refinement. This selective design directs computational resources toward volatile sub-sequences, thereby eliminating high-frequency noise without over-processing stable components. Defining \mathbf{X} as the multivariate input matrix of variables selected via Pearson Correlation Coefficient (PCC), the overall forecasting process is formulated as:

$$\hat{Y}_{\text{final}} = \sum_{j=1}^M \mathcal{H}_{\theta^*}(\mathcal{R}(C_j)) \quad (15)$$

Eq. (15) mathematically formalizes the final aggregation phase, grounded in the principle of superposition. Here, \hat{Y}_{final} denotes the ultimate forecasted wind power series. The summation operator \sum reconstructs the global signal by aggregating the individual predictions of M sub-components. Inside the summation, C_j represents the j -th component derived from the SE-based reconstruction. The operator $\mathcal{R}(\cdot)$ signifies the hierarchical refinement process (applying EWT denoising only to high-complexity modes), while $\mathcal{H}_{\theta^*}(\cdot)$ represents the mapping function of the hybrid BiTCN-BiLSTM-Attention predictor, parameterized by the SSA-optimized global solution θ^* .

To provide a rigorous algorithmic description of the implementation process, the detailed training strategy of MSAF-Net is formally summarized in Algorithm 1.

Algorithm 1: Training strategy of the proposed MSAF-Net framework

Require: Historical wind power series X , Meteorological dataset M , Max iterations T_{max} , Population size N
Ensure: Final wind power prediction \hat{Y}_{final}
1: Step 1: Feature Selection and Decomposition

 2: Calculate Pearson Correlation Coefficient (PCC) between X and variables in M

 3: $\mathcal{D} \leftarrow$ Select variables with significant correlation ($p < 0.05$)

 4: IMFs, $R \leftarrow$ CEEMDAN(\mathcal{D})

▷ Decompose inputs

5: Step 2: Entropy-Based Reconstruction (SE)

 6: Initialize $SE_{list} \leftarrow \emptyset$

 7: **for** each component c_k in IMFs **do**

 8: $SE_k \leftarrow$ SampEn(c_k)

 9: Append SE_k to SE_{list}

 10: **end for**

 11: Merge IMFs and Residual R based on SE values $\rightarrow \{C_1, C_2, \dots, C_m\}$

 12: $C_{high} \leftarrow$ argmax $_C(SE)$

▷ Identify highest complexity component

13: Step 3: Hierarchical Signal Refinement (EWT)

 14: $Sub_modes \leftarrow$ EWT(C_{high})

▷ Denoise high-frequency component

 15: $\mathcal{F} \leftarrow \{Sub_modes, C_{others}\}$

▷ Construct refined feature set

16: Step 4: SSA-Optimized Hybrid Prediction

 17: Initialize Sparrow population P (representing hyperparameters θ)

 18: **while** $t < T_{max}$ **do**

 19: **for** each sparrow $i = 1$ to N **do**

 20: Extract θ_i from position P_i

 21: Construct **BiTCN-BiLSTM-Attention** model \mathcal{M}_i with θ_i

 22: Calculate Fitness: $F_i \leftarrow$ RMSE $_{val}(\mathcal{M}_i(\mathcal{F}))$

 23: **end for**

 24: Update positions of Producers, Scroungers, and Scouts to minimize F

 25: $\theta_{best} \leftarrow$ Position with global minimum fitness

 26: $t \leftarrow t + 1$

 27: **end while**
28: Step 5: Final Aggregation

 29: Build optimal predictor \mathcal{H} utilizing θ_{best}

 30: $\hat{y}_{parts} \leftarrow \mathcal{H}.\text{predict}(\mathcal{F})$

▷ Forecast each component

 31: $\hat{Y}_{final} \leftarrow \sum \hat{y}_{parts}$

▷ Aggregate results

 32: **return** \hat{Y}_{final}

The implementation of the proposed framework, illustrated in the workflow of Fig. 4, begins with data preprocessing and decomposition to handle high-dimensional and non-stationary input. First, PCC analyzes the linear association between exogenous meteorological factors and wind power generation. By establishing a statistical significance threshold, the model retains only the variables exhibiting strong correlations to construct the input feature set X , minimizing redundant information that could compromise convergence. Following feature selection, the CEEMDAN algorithm decomposes both historical wind power series and the selected meteorological variables. As the first stage of the hierarchical strategy shown in Fig. 4, CEEMDAN resolves the complex raw signals into a series of IMF and a residual trend, reducing mode mixing and revealing multi-scale oscillatory characteristics within the data.

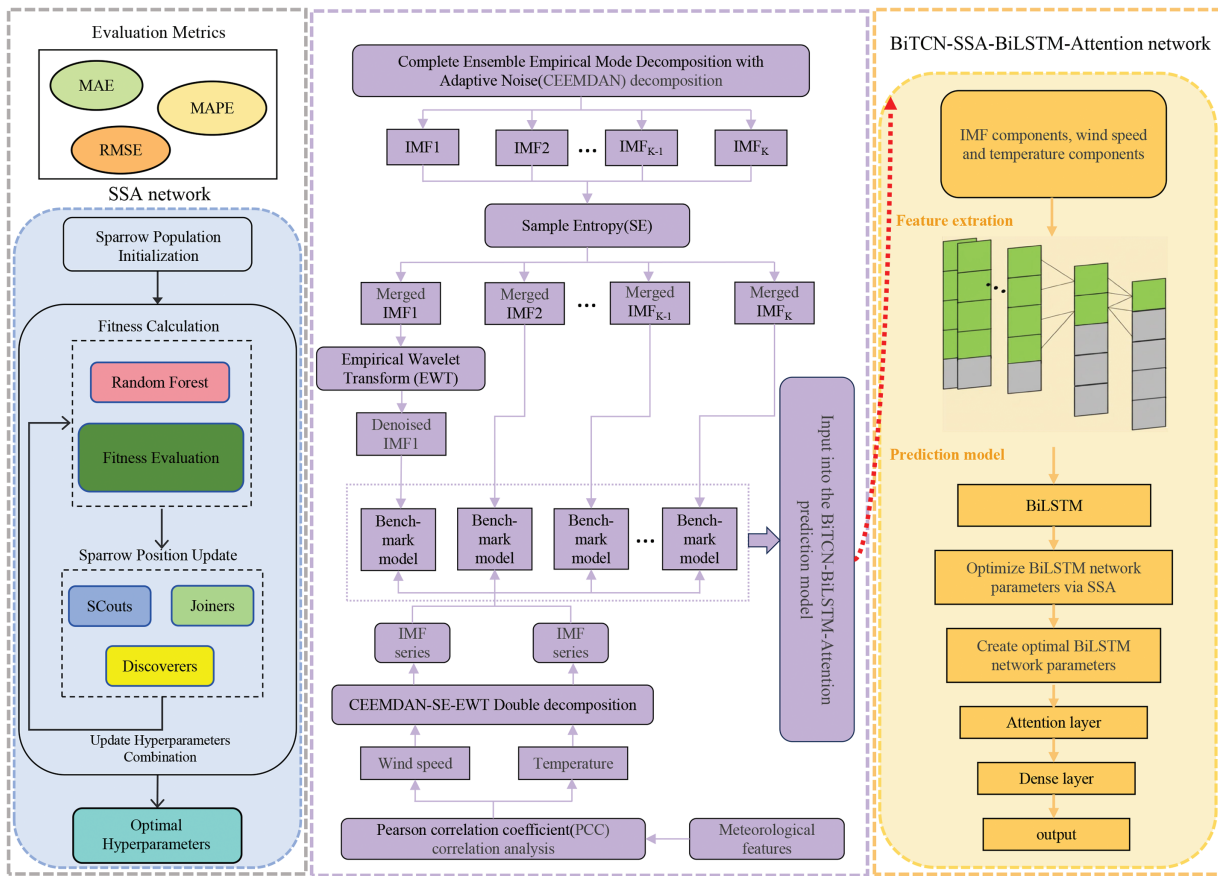


Figure 4: The general framework of the proposed method

To mitigate computational challenges and error accumulation risks associated with numerous decomposed modes, the framework implements a complexity-based reconstruction strategy. SE acts as a metric to assess the regularity and stochasticity of each extracted IMF. Based on entropy values, components with similar dynamic complexities merge into a compact set of reconstructed features $\{C_1, C_2, \dots, C_M\}$. This step, depicted in the Reconstruction module in Fig. 4, reduces the dimensionality of the prediction task while preserving essential signal information. At this juncture, the framework applies hierarchical signal refinement: EWT targets only the reconstructed component with the highest complexity (C_{high}), which contains the most severe high-frequency noise. This targeted secondary decomposition separates stochastic noise from deterministic high-frequency patterns without over-processing smoother low-frequency components, balancing noise reduction fidelity with computational efficiency.

Upon determining the refined feature matrix, the process moves to the hybrid forecasting engine, which forms the core of the MSAF-Net architecture shown in Fig. 4. The refined sub-sequences enter a BiTCN. By employing dilated causal convolutions within a bidirectional architecture, the BiTCN extracts local multi-scale features and spatial correlations from both forward and backward contexts, overcoming the receptive field limitations of standard convolutions. These high-level feature representations propagate to the BiLSTM network, which models long-term bidirectional temporal dependencies. To address the limitations of manual parameter tuning, the SSA serves as an optimization layer. As illustrated in the SSA Network block in Fig. 4, the SSA iteratively updates the positions of producers, scroungers, and scouts to identify the global optimal hyperparameter configuration θ^* (e.g., learning rate, number of hidden units) for the BiLSTM,

minimizing the validation RMSE. Furthermore, an Attention Mechanism embedded after the BiLSTM layer dynamically assigns probability weights to hidden states, prioritizing the contribution of critical time steps to the final prediction.

In the final workflow phase, the optimal hybrid predictor \mathcal{H}_{θ^*} generates independent forecasts for each refined component of the feature set. Linear aggregation of these partial predictions reconstructs the final wind power output \hat{Y}_{final} , reversing the decomposition process described in Eq. (15). This end-to-end procedure, spanning from correlation analysis to ensemble aggregation, enables the model to adapt to the varying volatilities of wind power data. Fig. 5 demonstrates the efficacy of the framework, presenting a comparative analysis of prediction results against actual power output and confirming the model's capability to track both rapid fluctuations and long-term trends.

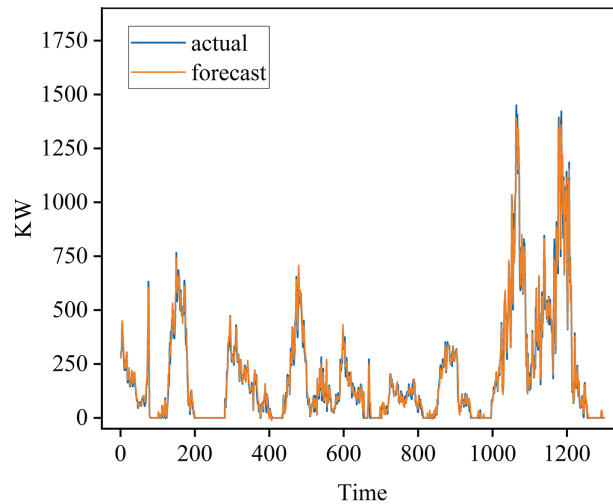


Figure 5: Comparison between the actual and forecast value

4 Experiments and Analysis

4.1 Data Description and Preprocessing

The experimental validation of the proposed MSAF-Net is conducted using real-world operational data collected from the La Haute Borne wind farm in France (sourced from the Engie Open Data platform), which is a widely recognized benchmark dataset. The wind farm comprises four turbines with a total installed capacity of 2050 kW. The dataset encompasses a comprehensive set of parameters sampled at 10-min intervals, including active power output, blade pitch angle, wind speed, wind direction, temperature, humidity, and atmospheric pressure.

To rigorously evaluate the robustness of the model, this study specifically utilizes a dataset recorded during the summer season (June to August). Unlike winter months, which are often characterized by consistent prevailing winds, the summer period exhibits significant thermal instability and rapid, stochastic wind speed fluctuations due to intense atmospheric convection. These characteristics result in a time series with stronger non-stationarity and higher complexity, providing an ideal “stress test” scenario. Validating the model on this high-volatility subset ensures that the reported performance improvements demonstrate the model's genuine capability to handle dynamic regimes inherent in renewable energy systems.

To ensure data quality and accelerate model convergence, a physics-informed preprocessing pipeline was applied prior to model training. First, a data cleaning strategy based on the physical characteristics of the turbine (Cut-in Speed $v_{cut} \approx 3.5$ m/s) was implemented to distinguish between natural calm conditions

and turbine faults: Samples where wind speed $v < 3.5$ m/s and Power $P \leq 0$ represent natural calm states and were retained (with negative power corrected to 0). Samples where wind speed $v \geq 3.5$ m/s but Power $P \leq 0$ indicate turbine faults or downtime. These were identified as outliers and removed to prevent the model from learning incorrect mapping relationships.

Missing values resulting from this cleaning were filled using linear interpolation for gaps shorter than one hour. Subsequently, to eliminate dimensional differences between multivariate features, all input variables were normalized to the $[0, 1]$ range using Min-Max scaling:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (16)$$

where x is the original value, and x_{min} and x_{max} are the minimum and maximum values of the feature column, respectively. Finally, the preprocessed dataset is partitioned into training, validation, and test sets in an 8:1:1 ratio to facilitate rigorous performance evaluation without data leakage.

4.2 Experimental Setup

The hardware configuration consists of an AMD Ryzen 7 7700 processor as the CPU and an NVIDIA GeForce RTX 5060 Ti as the GPU. Experiments were conducted using MATLAB R2024a, which offers a robust open-source community and high extensibility. The short-term wind power prediction network includes an initialization module and a hyperparameter optimization module. The optimization process addresses two categories of parameters: fixed architectural parameters and hyperparameters optimized with SSA.

To maintain training stability, standard hyperparameters were fixed based on preliminary trials. The Adam optimizer was selected for its adaptive learning rate capabilities, configured with a batch size of 64 and a maximum of 200 epochs. To prevent overfitting, a dropout rate of 0.3 was applied after the convolutional and LSTM layers. For the BiTCN module, a causal dilated convolution structure was adopted with 32 filters and a kernel size of 3 to extract multi-scale temporal features efficiently.

Unlike benchmark models that utilize static empirical settings, MSAF-Net employs SSA to dynamically search for optimal hyperparameters of the BiLSTM predictor. As detailed in Table 2, the search space covers the number of hidden units (N_{hidden}), the initial learning rate (α), and the L2 regularization factor (λ).

Table 2: Detailed hyperparameter settings: fixed configurations and SSA optimization ranges

Category	Parameter	Setting/Search range	Final value
Fixed config	Optimizer	Adam	–
	Batch size	64	–
	Max epochs	200	–
	BiTCN Filters	32	–
SSA optimized	Hidden units (N_{hidden})	Range: [10, 200]	128
	Learning rate (α)	Range: [0.001, 0.01]	0.0042
	L2 regularization (λ)	Range: [10^{-6} , 10^{-2}]	0.0015
SSA algorithm	Population size	20	–
	Max iterations	30	–

The SSA population size was set to **20**, with a maximum of **30 iterations**. To verify the sensitivity of model performance to population size, we conducted tests with sizes $N = \{10, 20, 30, 50\}$. Results indicated that while $N = 50$ slightly accelerated early convergence (as illustrated in Fig. 6), the final RMSE improvement was negligible ($<0.2\%$) compared to $N = 20$, but with double the computational cost. Therefore, a population of 20 was selected to balance optimization capability with efficiency. The specific search bounds and the final optimized values are listed in Table 2.

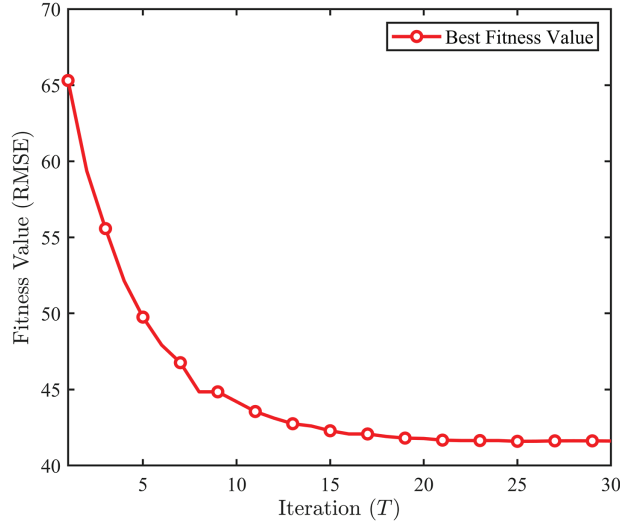


Figure 6: Convergence curve of the SSA optimization process

4.3 Evaluation Metrics

This study employs three common evaluation metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) [39]. These metrics are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (18)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_{max}} \right| \quad (19)$$

where y_i represents the actual wind power generation value at time i , \hat{y}_i denotes the predicted wind power generation value at time i , N indicates the number of data points. Due to the presence of zero values in the data, y_{max} is used as a variant of MAPE. Lower values of these evaluation metrics indicate better performance of the prediction method. This study also applies the Percentage Improvement to evaluate the enhancement of the proposed prediction method compared to other methods [40,41]. The Percentage Improvement is calculated as follows:

$$P_{MAE} = \frac{MAE_1 - MAE_2}{MAE_1} * 100\% \quad (20)$$

$$P_{RMSE} = \frac{RMSE_1 - RMSE_2}{RMSE_1} * 100\% \quad (21)$$

$$P_{MAPE} = \frac{MAPE_1 - MAPE_2}{MAPE_1} * 100\% \quad (22)$$

4.4 Discussion of Results

4.4.1 Feature Correlation Analysis

Prior to model training, the Pearson Correlation Coefficient (PCC) was employed to quantitatively assess the relevance of meteorological variables, thereby reducing computational complexity and eliminating redundant inputs. The significance level (p -value) was first examined to verify the linear association. As presented in Table 3, all calculated p -values (in parentheses) are less than 0.05, rejecting the null hypothesis and confirming statistically significant relationships.

Table 3: Pearson correlation coefficients between meteorological factors and wind power

Meteorological-factors	Wind speed	Wind direction	Temperature	Yaw angle	Humidity
Pearson Coefficient	0.97	0.19	0.67	0.18	0.26
Relevance	Strongly Related	Weakly Related	Strongly Related	Weakly Related	Weakly Related

The analysis reveals distinct correlation patterns. Wind speed and ambient temperature exhibit absolute PCC values of 0.97 and 0.67, respectively, indicating strong positive correlations with power generation. Conversely, wind direction, humidity, and yaw angle show weak associations ($|r| < 0.3$). Consequently, to optimize the input feature space, only wind speed and ambient temperature were selected as exogenous variables for the forecasting model. The PCC calculation is governed by Eq. (23):

$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (23)$$

4.4.2 Decomposition and Reconstruction Analysis

Wind power generation is inherently intermittent, a direct consequence of fluctuating meteorological conditions and topographical effects. To effectively model such a non-stationary time series, it is crucial to disentangle its constituent dynamics across multiple time scales. To this end, the raw wind power data was subjected to the CEEMDAN algorithm. This process decomposes the signal into 12 distinct IMF and a residual, as illustrated in Fig. 7. These components are naturally ordered by frequency, ranging from high-frequency, small-scale fluctuations (IMF1) to low-frequency, long-term trends (IMF12).

While this decomposition provides a high-resolution view, utilizing all 12 IMF directly can introduce excessive model complexity and computational burden. To create a more parsimonious and robust feature set, we implemented a principled reconstruction strategy based on component complexity. First, SE was employed to provide a quantitative measure of the regularity and predictability of each IMF. The resulting SE values, listed in Fig. 8, empirically confirm the expected decrease in complexity from the high-frequency to the low-frequency modes. Based on this objective analysis, IMFs exhibiting proximate SE values—indicating similar dynamic characteristics—were merged. This consolidation yielded a reconstructed set of seven new components: IMF3 and IMF4 were combined into New IMF3; IMF6 and IMF7 into New IMF5; IMF8 and IMF9 into New IMF6; and the three lowest-frequency modes (IMF10-12) into New IMF7. Through this data-driven merging process, the initial 12 IMF were thoughtfully reconstructed into a more manageable and meaningful set of seven components, each representing a distinct dynamic scale for subsequent forecasting.

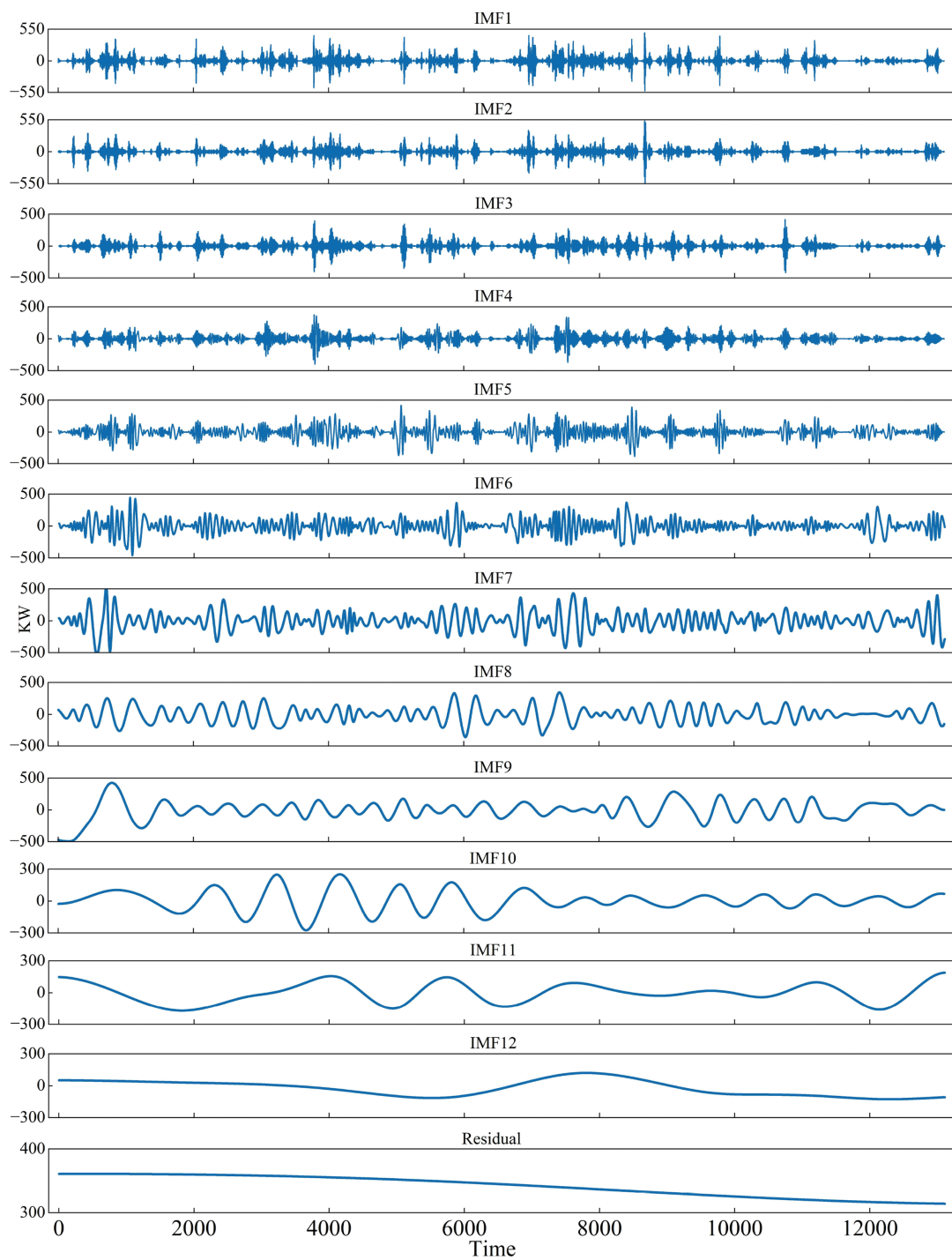


Figure 7: The results of CEEMDAN decomposition

The New IMF1 component, which has the highest complexity, is further decomposed using EWT, as illustrated in Fig. 9. The SE values of IMF merged according to their similar sample entropy values are summarized in Table 4. It can be observed that the sample entropy values of the decomposed sub-sequences exhibit lower complexity. The decomposed sub-sequences (shown in Fig. 10) are then input into the prediction model along with wind speed and temperature sequences.

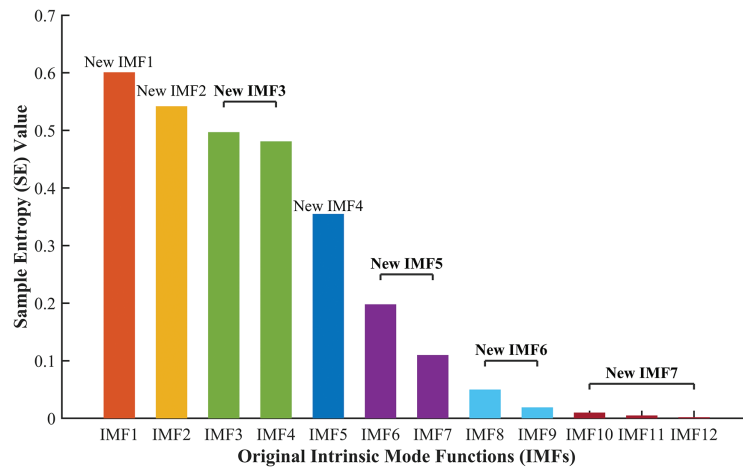


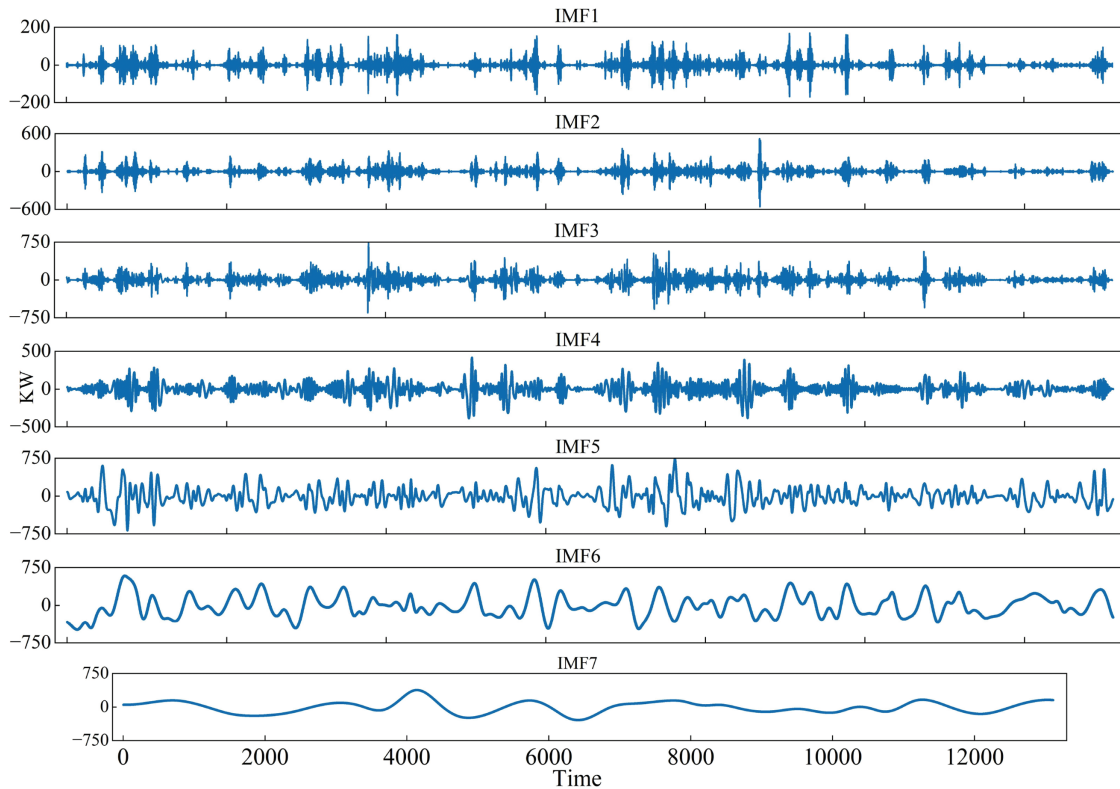
Figure 8: Wind power component sample entropy value



Figure 9: EWT decomposition results

Table 4: The SE values of IMF merged based on their similar sample entropy values

Serial number	SE value	Serial number	SE value
IMF1	0.601	IMF5	0.134
IMF2	0.542	IMF6	0.042
IMF3	0.452	IMF7	0.006
IMF4	0.355		

**Figure 10:** The subsequence of the final input

4.4.3 Comparison with Benchmark Methods

To comprehensively evaluate the performance of MSAF-Net and quantify the specific contribution of each module, five representative benchmark models were selected to form a systematic ablation study.

First, the single deep learning model LSTM was chosen to establish a fundamental performance baseline for processing non-stationary wind power data without auxiliary decomposition.

Second, a hybrid predictor without decomposition, BiTCN-BiLSTM, was included to isolate the benefits of the signal decomposition strategy, verifying whether the structural integration of convolutional and recurrent networks alone is sufficient to capture complex fluctuations.

Third, three decomposition-based variants were selected to validate specific components of the proposed framework: (1) EMD-BiTCN-BiLSTM is compared against the proposed CEEMDAN-based model to justify the superiority of CEEMDAN in mitigating mode mixing issues inherent in EMD; (2) CEEMDAN-EWT-LSTM employs the same hierarchical decomposition as MSAF-Net but uses a simpler predictor,

thereby isolating and validating the architectural advantage of the BiTCN-BiLSTM network over standard LSTM; and (3) CEEMDAN-BiTCN-BiLSTM serves as a critical ablation baseline to specifically quantify the performance gains attributed to the proposed secondary denoising (SE-EWT) and adaptive SSA optimization, demonstrating the necessity of these modules for eliminating residual noise and avoiding local optima.

To establish a consistent baseline, all benchmark models were trained using the Adam optimizer for a maximum of 200 epochs with a dropout rate of 0.3. Input and output dimensions were standardized to 3 and 1, respectively, with a time step of 7 determined through trial-and-error. For decomposition-based methods, white noise was added 200 times. As shown in Table 2, the proposed MSAF-Net incorporated SSA to optimize various hyperparameters, fully leveraging the potential of the deep learning architecture.

A comparison of different methods on the dataset is presented in Table 5, and the percentage improvements are summarized in Table 6. Based on these results, a clear hierarchical path to achieving superior predictive performance is evident. Initially, the benefit of combined models is apparent; the MAE, RMSE, and MAPE values of the BiTCN-BiLSTM method are consistently lower than those of the single LSTM prediction method. Furthermore, the most significant performance leap is achieved by addressing the inherent volatility of the raw wind data. Due to the high non-stationarity of the original time series, methods that do not employ decomposition struggle. Consequently, the hybrid decomposition methods (EMD-BiTCN-BiLSTM, CEEMDAN-BiTCN-BiLSTM, CEEMDAN-EWT-LSTM, and the proposed approach) demonstrate decisively better MAPE values than all non-decomposition models. This confirms that decomposing the signal into more stable sub-series is a critical step for improving prediction quality.

Table 5: Performance accuracy of different forecasting methods

	MAE	RMSE	MAPE
LSTM	82.47	136.27	3.92
BiTCN-BiLSTM	74.81	114.22	3.41
EMD-BiTCN-BiLSTM	52.37	81.21	2.41
CEEMDAN-BiTCN-BiLSTM	42.63	68.78	1.92
CEEMDAN-EWT-LSTM	36.20	64.12	1.77
MSAF-Net	26.67	41.59	1.36

Table 6: Percentage improvement of proposed method vs. other benchmarking methods

	P_{MAE}	P_{RMSE}	P_{MAPE}
MSAF-Net vs. LSTM	67.66%	69.48%	65.31%
MSAF-Net vs. BiTCN-BiLSTM	64.35%	63.59%	60.12%
MSAF-Net vs. EMD-BiTCN-BiLSTM	49.07%	48.79%	53.26%
MSAF-Net vs. CEEMDAN-BiTCN-BiLSTM	37.44%	39.53%	29.17%
MSAF-Net vs. CEEMDAN-EWT-LSTM	26.32%	35.14%	23.16%

Within the family of hybrid decomposition methods, the choice of technique proves crucial. CEEMDAN demonstrates superior decomposition capability over its predecessors. For example, the MAPE value of CEEMDAN-BiTCN-BiLSTM is 1.92%, which is markedly lower than the 2.41% achieved by its EMD-BiTCN-BiLSTM counterpart. This aligns with established conclusions regarding the effectiveness of adaptive noise in mitigating mode mixing [27]. In addition, the study confirms the value of a two-stage decomposition

strategy. The MAPE value of the CEEMDAN-EWT-LSTM model is 1.77%, which outperforms single-stage decomposition models, demonstrating that smoothing and denoising the high-frequency IMF1 using EWT technology effectively enhances prediction accuracy.

This hierarchical validation culminates in the proposed MSAF-Net method, which shows the highest accuracy compared to all aforementioned methods. Its success is rooted in the synergistic integration of every component: classifying and reconstructing IMFs based on sample entropy reduces experimental complexity, while denoising the first IMF using EWT mitigates the negative effects of randomness. Building on this, the combined BiTCN-BiLSTM prediction model outperforms single prediction models, a performance that is significantly enhanced by the SSA optimizing the BiLSTM hyperparameters. Moreover, the attention mechanism strengthens the model's ability to analyze the impact of meteorological factors, further refining prediction precision. By fully integrating these advantages, the proposed MSAF-Net achieves a MAPE of 1.36%, standing as an effective tool for ultra-short-term wind power prediction. For a clearer visual comparison of the performance across different models, the prediction curves are presented in Fig. 11.

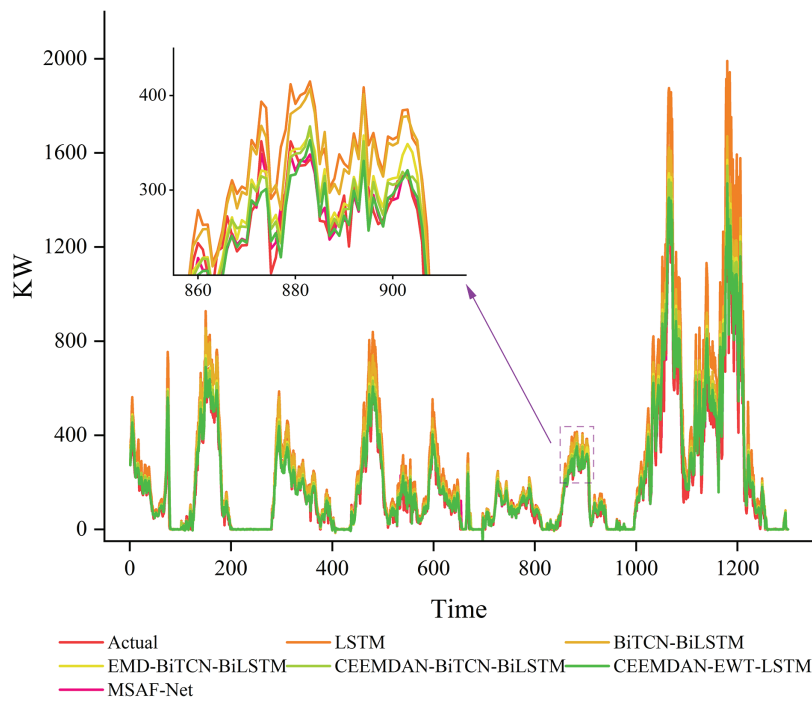


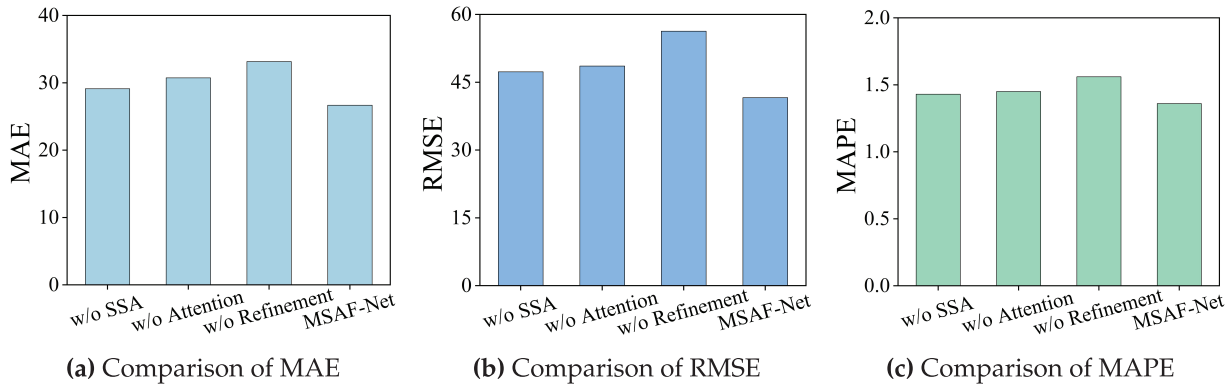
Figure 11: Prediction curves of various models

4.4.4 Ablation Study

To systematically investigate the internal mechanisms and verify the criticality of individual components within the MSAF-Net framework, an ablation study was conducted using three specific variants. For clarity in the subsequent analysis, the mappings between the full architecture names and their abbreviated variants are defined as follows: **w/o Refinement** corresponds to CEEMDAN-BiTCN-SSA-BiLSTM-Attention, which excludes the hierarchical signal refinement strategy; **w/o SSA** denotes CEEMDAN-SE-EWT-BiTCN-BiLSTM-Attention, where the intelligent optimization is replaced by empirical manual tuning; and **w/o Attention** refers to CEEMDAN-SE-EWT-BiTCN-SSA-BiLSTM, which removes the temporal attention mechanism. The quantitative performance metrics of these variants against the complete MSAF-Net are listed in Table 7, and the visual comparisons are presented in Fig. 12.

Table 7: Quantitative comparison of MSAF-Net and its ablation variants

	MAE	RMSE	MAPE
w/o SSA	29.12	47.31	1.43
w/o Attention	30.74	48.58	1.45
w/o Refinement	33.16	56.29	1.56
MSAF-Net	26.67	41.59	1.36

**Figure 12:** Performance evaluation of component contributions via error metrics (MAE, RMSE, and MAPE)

A comprehensive analysis of the ablation results reveals the distinct functional contribution of each module to the overall prediction accuracy. The most prominent observation is that the **w/o Refinement** variant exhibited the severest deterioration in performance across all metrics, with MAE and RMSE rising significantly compared to the proposed model. This substantial gap highlights the vulnerability of deep learning predictors to the intrinsic non-stationarity of raw wind data. Without the SE-based reconstruction and EWT secondary denoising, the predictor struggles to distinguish between meaningful trends and high-frequency stochastic noise, leading to instability. Following this, the performance decline in the **w/o SSA** model underscores the limitations of static hyperparameter settings. While the deep learning structure remains intact, the lack of adaptive optimization forces the network to operate in a suboptimal state, failing to fully capture the complex mapping relationships within the dataset. In contrast, SSA effectively navigates the hyperparameter space to prevent local optima, ensuring the model's complexity aligns with the data characteristics.

Furthermore, the comparison regarding the **w/o Attention** variant provides critical insights into model interpretability. Although its overall error is lower than the other two variants, it exhibits increased lag during power ramp events. This deficiency highlights that the Attention Mechanism serves as more than a performance enhancer; it acts as an interpretive lens that breaks the “black box” nature of deep learning. By dynamically assigning higher probability weights to the most recent or salient time steps, the mechanism effectively identifies critical “turning points” in the wind power sequence. In contrast, the **w/o Attention** variant treats all historical time steps with equal importance, failing to prioritize these decisive features when the wind regime changes abruptly. Consequently, the complete MSAF-Net achieves the lowest error metrics by synergizing these components: the refinement strategy ensures high-quality input, SSA guarantees optimal model configuration, and the attention mechanism enhances both the sensitivity to dynamic changes and the interpretability of the prediction process.

4.4.5 Computational Complexity Analysis

As presented in Table 8, simple models like LSTM require minimal training time (45.20 s) due to their shallow architecture and lack of signal decomposition. In contrast, the decomposition-based methods (e.g., CEEMDAN-EWT-LSTM) incur higher computational costs because the decomposition process must be performed on the training set. For the proposed MSAF-Net, the training phase is indeed the most computationally intensive (2450.85 s), primarily driven by the iterative search process of the SSA, which evaluates the fitness of 20 sparrows over 30 iterations.

Table 8: Comparison of computational efficiency (Training vs. Inference) on the test set

Model	Training time (s)	Inference time (ms/Sample)
LSTM	45.20	0.52
BiTCN-BiLSTM	186.45	2.10
CEEMDAN-EWT-LSTM	712.30	3.85
MSAF-Net	2450.85	4.55

However, for practical engineering applications such as real-time grid dispatching, the critical metric is the Online Inference Time. Once the model is trained and the optimal hyperparameters (θ^*) are fixed, the SSA optimization loop is no longer required during the prediction phase. Consequently, the inference process only involves signal decomposition and a forward pass of the deep neural network. As shown in Table 8, the average inference time of MSAF-Net is approximately 4.55 ms per sample. Given that ultra-short-term wind power dispatching typically operates on a 15-min (900,000 ms) interval, the inference latency of MSAF-Net is negligible.

This analysis confirms that MSAF-Net achieves a favorable trade-off: it accepts a higher one-time offline training cost to secure superior predictive accuracy (lowest MAPE) and efficient real-time deployment capabilities.

5 Conclusion

Accurate ultra-short-term wind power forecasting is a prerequisite for safe renewable energy integration. To mitigate stochastic volatility and non-stationarity, this paper proposes MSAF-Net, a hybrid framework synergizing hierarchical signal refinement (CEEMDAN-SE-EWT) with an optimized deep learning predictor (BiTCN-SSA-BiLSTM-Attention). Validation on the La Haute Borne benchmark demonstrates that the model outperforms traditional decomposition-based methods, achieving a MAPE of 1.36%. Beyond numerical superiority, the work offers substantial practical implications. The model's capacity to track rapid fluctuations allows grid operators to anticipate ramp events with precision, enhancing frequency stability and reducing reserve capacity costs.

Despite superior predictive performance, this study identifies limitations. First, the SSA optimization loop increases offline training computational overhead compared to non-optimized models. While online inference remains efficient for real-time dispatch, high training costs may challenge scenarios requiring frequent retraining. Second, while the model underwent stress-testing with high-volatility summer data, its adaptability to extreme weather (e.g., typhoons, icing) and diverse geographical terrains requires verification on a broader scale.

Future research will address these constraints by investigating surrogate-assisted evolutionary algorithms to accelerate optimization. Furthermore, incorporating physical constraints, such as Numerical

Weather Prediction physics, into the loss function will improve physical consistency. Research will also extend the framework from point forecasting to multi-step probabilistic forecasting, quantifying uncertainty to support risk-aware decision-making.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by Yunnan Fundamental Research Projects (202401CF070073).

Author Contributions: The authors confirm contributions to the study as follows: Xiaolan Li: Writing—review and editing, Methodology, Funding acquisition, Data curation; Jinyu Shen: Writing—original draft, Software, Methodology, Conceptualization; Jinhua Liang: Data curation; Yanting Wang: Software. All authors contributed to the discussion of results, reviewed the manuscript, and approved the final version for submission. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data will be made available on request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zhang F, Li N, Li L, Wang S, Du C. A local semi-supervised ensemble learning strategy for the data-driven soft sensor of the power prediction in wind power generation. *Fuel*. 2023;333(11):126435. doi:10.1016/j.fuel.2022.126435.
2. Guo Y, Wang H, Lian J. Review of integrated installation technologies for offshore wind turbines: current progress and future development trends. *Energy Convers Manag*. 2022;255(4):115319. doi:10.1016/j.enconman.2022.115319.
3. Teleke S, Baran ME, Bhattacharya S, Huang AQ. Optimal control of battery energy storage for wind farm dispatching. *IEEE Trans Energy Convers*. 2010;25(3):787–94. doi:10.1109/tec.2010.2041550.
4. Farah S, Humaira N, Aneela Z, Steffen E. Short-term multi-hour ahead country-wide wind power prediction for Germany using gated recurrent unit deep learning. *Renew Sustain Energy Rev*. 2022;167(2):112700. doi:10.1016/j.rser.2022.112700.
5. Kong X, Ma L, Wang C, Guo S, Abdelbaky MA, Liu X, et al. Large-scale wind farm control using distributed economic model predictive scheme. *Renew Energy*. 2022;181(3/4):581–91. doi:10.1016/j.solener.2024.112798.
6. Abdelbaky MA, Liu X, Jiang D. Design and implementation of partial offline fuzzy model-predictive pitch controller for large-scale wind-turbines. *Renew Energy*. 2020;145:981–96. doi:10.1016/j.renene.2019.05.074.
7. Jin H, Shi L, Chen X, Qian B, Yang B, Jin H. Probabilistic wind power forecasting using selective ensemble of finite mixture Gaussian process regression models. *Renew Energy*. 2021;174(3):1–18. doi:10.1016/j.renene.2021.04.028.
8. Aly HH. A hybrid optimized model of adaptive neuro-fuzzy inference system, recurrent Kalman filter and neuro-wavelet for wind power forecasting driven by DFIG. *Energy*. 2022;239:122367. doi:10.1016/j.energy.2021.122367.
9. Xiong J, Peng T, Tao Z, Zhang C, Song S, Nazir MS. A dual-scale deep learning model based on ELM-BiLSTM and improved reptile search algorithm for wind power prediction. *Energy*. 2023;266(2):126419. doi:10.1016/j.energy.2022.126419.
10. Du P, Wang J, Yang W, Niu T. A novel hybrid model for short-term wind power forecasting. *Appl Soft Comput*. 2019;80(1–2):93–106. doi:10.1016/j.asoc.2019.03.035.
11. Shahid F, Zameer A, Muneeb M. A novel genetic LSTM model for wind power forecast. *Energy*. 2021;223(1):120069. doi:10.1016/j.energy.2021.120069.
12. Mohsen S, Ghoneim SS, Alzaidi MS, Alzahrani A, Hassan AMA. Classification of electroencephalogram signals using LSTM and SVM based on fast Walsh-Hadamard transform. *Comput Mater Contin*. 2023;75(3):5271–86. doi:10.32604/cmc.2023.038758.
13. Joseph LP, Deo RC, Prasad R, Salcedo-Sanz S, Raj N, Soar J. Near real-time wind speed forecast model with bidirectional LSTM networks. *Renew Energy*. 2023;204(7):39–58. doi:10.1016/j.renene.2022.12.123.

14. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* 2005;18(5–6):602–10. doi:10.1016/j.neunet.2005.06.042.
15. Fan GF, Li JW, Peng LL, Huang HP, Hong WC. The bi-long short-term memory based on multiscale and mesoscale feature extraction for electric load forecasting. *Appl Soft Comput.* 2024;162:111853. doi:10.1016/j.asoc.2024.111853.
16. Peng S, Zhu J, Wu T, Yuan C, Cang J, Zhang K, et al. Prediction of wind and PV power by fusing the multi-stage feature extraction and a PSO-BiLSTM model. *Energy.* 2024;298(8):131345. doi:10.1016/j.energy.2024.131345.
17. Li N, Xu W, Zeng Q, Ren Y, Ma W, Tan K. A hybrid WOA-CNN-BiLSTM framework with enhanced accuracy for low-voltage shunt capacitor remaining life prediction in power systems. *Energy.* 2025;326(20):136183. doi:10.1016/j.energy.2025.136183.
18. Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst Sci Control Eng.* 2020;8(1):22–34. doi:10.1080/21642583.2019.1708830.
19. Limouni T, Yaagoubi R, Bouziane K, Guissi K, Baali EH. Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. *Renew Energy.* 2023;205:1010–24. doi:10.1016/j.renene.2023.01.118.
20. Chen G, Li X, Zhang R, Zhang H, Han J, Zhang T. Short-term offshore wind speed prediction model based on VMD-GDPSO-TCN-BiLSTM. *Ocean Eng.* 2025;341(4):122518. doi:10.1016/j.oceaneng.2025.122518.
21. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing.* 2019;337(5):325–38. doi:10.1016/j.neucom.2019.01.078.
22. Mu G, Yang M, Wang D, Yan G, Qi Y. Spatial dispersion of wind speeds and its influence on the forecasting error of wind power in a wind farm. *J Mod Power Syst Clean Energy.* 2016;4(2):265–74. doi:10.1007/s40565-015-0151-x.
23. Peng Z, Peng S, Fu L, Lu B, Tang J, Wang K, et al. A novel deep learning ensemble model with data denoising for short-term wind speed forecasting. *Energy Convers Manag.* 2020;207(1):112524. doi:10.1016/j.enconman.2020.112524.
24. Yang G, Yuan E, Wu W. Predicting the long-term CO₂ concentration in classrooms based on the BO-EMD-LSTM model. *Build Environ.* 2022;224(1):109568. doi:10.1016/j.buildenv.2022.109568.
25. Chen Y, Dong Z, Wang Y, Su J, Han Z, Zhou D, et al. Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history. *Energy Convers Manag.* 2021;227(4):113559. doi:10.1016/j.enconman.2020.113559.
26. Fang P, Fu W, Wang K, Xiong D, Zhang K. A composite architecture coupling outlier correction, EWT, nonlinear Volterra multi-model fusion with multi-objective optimization for short-term wind speed forecasting. *Appl Energy.* 2022;307(1):118191. doi:10.1016/j.apenergy.2021.118191.
27. Ren Y, Suganthan PN, Srikanth N. A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods. *IEEE Trans Sustain Energy.* 2014;6(1):236–44. doi:10.1109/tste.2014.2365580.
28. Lv P, Shu Y, Xu J, Wu Q. Modal decomposition-based hybrid model for stock index prediction. *Expert Syst Appl.* 2022;202(1):117252. doi:10.1016/j.eswa.2022.117252.
29. Lin Y, Lin Z, Liao Y, Li Y, Xu J, Yan Y. Forecasting the realized volatility of stock price index: a hybrid model integrating CEEMDAN and LSTM. *Expert Syst Appl.* 2022;206(4):117736. doi:10.1016/j.eswa.2022.117736.
30. Karijadi I, Chou SY, Dewabharata A. Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method. *Renew Energy.* 2023;218:119357. doi:10.1016/j.renene.2023.119357.
31. Li K, Huang W, Hu G, Li J. Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. *Energy Build.* 2023;279(84):112666. doi:10.1016/j.enbuild.2022.112666.
32. Su Y, Wang Z, Dong Z, Hua X, Ye T, Song Z, et al. Frequency-aware ultra-short-term wind power forecasting using CEEMDAN-VMD-SE and Transformer-GRU networks. *Energy.* 2025;338(5):138715. doi:10.1016/j.energy.2025.138715.
33. Zhou D, Liu Y, Wang X, Wang F, Jia Y. Combined ultra-short-term photovoltaic power prediction based on CEEMDAN decomposition and RIME optimized AM-TCN-BiLSTM. *Energy.* 2025;318(1):134847. doi:10.1016/j.energy.2025.134847.
34. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271. 2018.

35. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A Math Phys Eng Sci.* 1998;454(1971):903–95. doi:10.1098/rspa.1998.0193.
36. Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2011 May 22–27; Prague, Czech Republic. p. 4144–7.
37. Gilles J. Empirical wavelet transform. *IEEE Trans Signal Process.* 2013;61(16):3999–4010. doi:10.1109/tsp.2013.2265222.
38. Samal KKR, Panda AK, Babu KS, Das SK. Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites. *Urban Clim.* 2021;39:100943. doi:10.1016/j.uclim.2021.100943.
39. Hong T, Fan S. Probabilistic electric load forecasting: a tutorial review. *Int J Forecast.* 2016;32(3):914–38. doi:10.1016/j.ijforecast.2015.11.011.
40. Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. *Energy Convers Manag.* 2019;198(11):111799. doi:10.1016/j.enconman.2019.111799.
41. Makridakis S, Spiliotis E, Assimakopoulos V. The M4 competition: results, findings, conclusion and way forward. *Int J Forecast.* 2018;34(4):802–8. doi:10.1016/j.ijforecast.2018.06.001.