



OPEN Multi-scale wind speed prediction model based on improved escape algorithm for optimizing time-varying filtering empirical modal decomposition

Haili Zheng^{1,2}, Qian Wu^{1,2}, Xuanhan Lv¹, Peitong Zeng¹, Xiaoli Zhu¹ & YuPan Zeng¹✉

Wind energy's stochasticity and volatility challenge grid stability and dispatch reliability. To overcome limitations of existing decomposition-based forecasting methods, this paper proposes an Improved Escape Algorithm (IESC) incorporating chaotic mapping to optimize the hyperparameters of Time-Varying Filter Empirical Mode Decomposition (TVF-EMD), effectively mitigating mode mixing and enhancing non-stationary wind signal separation. Departing from uniform modeling, we employ a frequency-adaptive strategy: XLSTM captures high-frequency volatility, LSTM models medium-frequency transitions, and ELM rapidly processes low-frequency trends. Evaluated on a large-scale dataset, IESC outperforms standard ESC, GWO, and DE by 6.2%, 11.3%, and 8.4%, respectively. The proposed hybrid model demonstrates superior robustness, achieving a 29.8% lower 1-step MAE (0.5109) and a 65.6% higher 15-step R² (0.7685) compared to XLSTM alone. Crucially, error growth (1–15 steps) is contained within 12% and R² degradation is 35% slower. These results confirm that the method significantly enhances forecasting precision and effectively bridges multi-step accuracy with real-time dispatch needs, ensuring dynamic grid-demand matching and improved operational stability.

Keywords Improved escape algorithm, Multi-scale prediction model, Time-varying filter empirical mode decomposition, Wind speed prediction

In the global transition to clean energy, wind energy is a vital renewable resource—its large-scale deployment is key to addressing energy crises, cutting carbon emissions, and ensuring the reliable operation of modern power systems—including microgrids (MGs) and community multi-energy systems. As wind power capacity expands, its growing share of generation mixes is undermined by inherent intermittency and volatility, which disrupt microgrid day-ahead scheduling, reduce flexibility-oriented energy management efficiency: the day-ahead scheduling of integrated MGs, the efficiency of flexibility management in renewable communities, and the coordination between TSO and DSO in energy markets^{1–3}. Accurate wind speed prediction thus becomes critical to enhancing system stability, improving wind power grid integration, optimizing energy allocation, and supporting risk-averse decisions in renewable-rich networks⁴.

To enhance wind speed prediction performance, research on wind speed prediction has evolved from physical models to data-driven approaches. Early time-series models such as Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA)^{5,6} capture linear temporal correlations but fail to handle nonlinear, high-volatility wind sequences—often causing microgrid (MG) power imbalances¹.

With advancing AI, shallow machine learning models^{7–10} have been widely used—they integrate multi-source meteorological features to improve short-term prediction accuracy. However, their insensitivity to long-term temporal patterns and vanishing gradient issues in large-scale data processing^{11,12} limit applicability in flexibility market multi-step prediction³.

To address these limitations, deep learning models based on Recurrent Neural Networks (RNN)¹³, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs)¹⁴, utilize memory cells to capture long-term dependencies. These models outperform shallow models in multi-step prediction tasks. Nevertheless,

¹School of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang 524088, China. ²Haili Zheng and Qian Wu contributed equally to this work. ✉email: zengyupan@gdou.edu.cn

standalone deep models still struggle with wind volatility, particularly when combined with other uncertainties in MGs¹. This challenge has driven the development of hybrid models that integrate time-series decomposition with prediction.

Time-series decomposition reduces data complexity by separating wind sequences into trend and noise components. Common techniques include Empirical Mode Decomposition (EMD)¹⁵, Ensemble Empirical Mode Decomposition (EEMD)^{16,17}, and Variational Mode Decomposition (VMD)¹⁸. Among these, Time-Varying Filter Empirical Mode Decomposition (TVF-EMD)¹⁹ is notable for its time-varying filter, which enables precise separation of frequency components—a key feature for distinguishing short-term wind fluctuations from long-term trends. However, TVF-EMD performance depends on two parameters: B-spline curve order and bandwidth threshold²⁰.

Heuristic algorithms like Genetic Algorithm (GA)²¹, Particle Swarm Optimization (PSO)²², Differential Evolution (DE)²³, Grey Wolf Optimizer (GWO)²⁴ optimize these parameters but suffer from poor convergence or insufficient diversity. The Escape Algorithm (ESC) solves this via crowd escape dynamics²⁵, ESC has the advantages of superior convergence performance, higher search efficiency, and stronger output stability compared to other algorithms. But ESC lacks initial population diversity and uniform solution distribution—limiting its use for complex wind data. To address this limitation, we focus on optimizing the ESC algorithm, leveraging the improved version (IESC) to adjust the bandwidth threshold and B-spline order of TVF-EMD.

The optimized TVF-EMD then decomposes wind speed data into high-frequency noise signals, medium-frequency components, and low-frequency trend signals. The volatile, complex noise requires high-complexity models to avoid vanishing gradients, while the smoother trend is susceptible to overfitting with such models²⁶. Therefore, separate modeling is essential for accurate feature extraction²⁷. These are predicted separately using XLSTM²⁸, LSTM, and ELM, respectively. The final forecast is obtained by reconstructing all predicted components.

This study proposes a novel multi-scale wind speed prediction model with five core innovations:

- (1) Novel Improved Escape Optimization (IESC) enhances TVF-EMD parameter tuning.
- (2) Multi-scale model combines XLSTM, LSTM, and ELM for frequency-specific prediction.
- (3) IESC outperforms GA, PSO, and DE in convergence and decomposition efficiency.
- (4) Achieves low-error wind speed forecasting for 1–15 time steps.
- (5) Optimized TVF-EMD effectively separates noise and trend components.

In this study, the IESC is theoretically analysed in section “[Time-varying filtering empirical modal decomposition optimized via improved escape algorithm](#)”, along with an examination of TVF-EMD. The multi-scale prediction model is explored in section “[Multi-scale predictive modeling](#)”, evaluation metrics for decomposition and prediction models are discussed in section “[Model evaluation indicators](#)”, and arithmetic examples are analyzed in section “[Calculus analysis](#)”. The study is concluded in section “[Conclusion](#)”.

Time-varying filtering empirical modal decomposition optimized via improved escape algorithm

This section details the structure of the Improved Escape Algorithm (IESC), and the Time-Varying Filtering Empirical Mode Decomposition (TVF-EMD). In addition, the workflow of the improved algorithm is provided.

Improved escape algorithm

In this work we improve the ESC algorithm from the following four aspects:

- (1) Logistic chaotic mapping initialization.
- (2) Convergence-oriented Elite Pool and Grouping Strategy.
- (3) Introductory Mutation Strategy.
- (4) Optimal Attraction Strategy.

Logistic chaotic mapping initialization

Logistic chaos initialization ensures population uniformity, boosting global search²⁹ (Eq. 1).

$$x_{i+1} = \mu x_i (1 - x_i) \quad (1)$$

Where x_i is an individual under a certain dimension, and μ is a control parameter with the value of (0,4). The larger μ is, the higher the degree of system chaos.

Convergence-oriented elite pool and grouping strategy

During search, the population’s unawareness of safe exits leads to an assumed time-increasing discovery of exits (potential optima). This mechanism dynamically adjusts the Elite Pool via Eq. (2) to strengthen its guidance.

$$E_{t+1} = [E_t + t \cdot E_{factor} / T] \quad (2)$$

Where E_t : Elite Pool (exit) individuals at iteration t ; t : current iteration index; T : total iterations; E_{factor} : adaptive Elite Pool factor (per Eq. 3).

$$E_{factor} = E_j * rand_j \quad (3)$$

Where $rand_j$ is the random number generated between (0,1).

Post-iteration, portions of the Herding Group transition to the Calm Group, while Panic Group individuals cease random movement and join the Herding Group. Equation (4) dynamically regulates this transition to boost elite guidance on underperforming individuals.

$$c_{t+1} = c_t + \text{sigmoid}(5(t-1)/T) \cdot (t \cdot \exp(t/T)/\pi) \quad (4)$$

Where c_t : Elite Pool individual count at iteration t ; Sigmoid function serves as the smooth transition (Eq. 5).

$$S(x) = 1/(1 + \exp(-x)) \quad (5)$$

The transition from the Panic Group to the Herding Group follows a pattern analogous to the Calm Group. With increasing iterations, both groups exhibit a rising proportion of individuals moving toward superior solutions.

Introductory mutation strategy

During the exploratory phase, the convergence of the tri-group towards elites after each iteration resulted in a loss of diversity. This issue was mitigated through differentiated mutation strategies:

In the Panic Group, stochastic perturbations were intensified to enhance diversity, while guidance mechanisms directed low-fitness individuals in the Calm and Herding Groups towards elites.

The Calm Group employs the elite differential variant (DE/best/2)³⁰ to cull and replace its bottom 25% individuals, which harnesses elite-guided information to steer convergence, supplements perturbation vectors to sustain population diversity, and enhances group quality while steering toward the optimal solution defined in Eq. (6).

$$v_{i,j+1} = x_{best,j} + F \cdot (x_{r1,j} - x_{r2,j}) + F \cdot (x_{r3,j} - x_{r4,j}) \quad (6)$$

Where $x_{best,j}$: optimal individual in the calm group; $x_{r1,j}, x_{r2,j}, x_{r3,j}, x_{r4,j}$: four random individuals from the calm group; F : scaling factor.

Secondly, the Herding Group replaces the Calm Group's bottom 30% via DE/rand-to-best/1³¹, steering toward superior solutions and enhancing quality (Eq. 7). This strategy balances random diversity with elite traits, stabilizing the search process.

$$v_{i,j+1} = x_{i,j} + F \cdot (x_{best,j} - x_{i,j}) + F \cdot (x_{r1,j} - x_{r2,j}) \quad (7)$$

Where $x_{best,j}$: optimal in the slave group; $x_{r1,j}, x_{r2,j}$: two random individuals from the slave group; F : scaling factor.

Normal cloud variation³² replaces the Panic Group's bottom 25% to boost randomness. Multi-dimensional entropy En'_k is quantified via Eq. (8).

$$En'_k = En_k + He_k \cdot \epsilon_k, \epsilon_k \sim N(0, 1) \quad (8)$$

En_k : entropy (full-1(0.5) vector); He_k : hyper-entropy (full-1(0.1) vector). Cloud droplets are generated via dynamic entropy (Eq. 9).

$$x_{k,m} \sim N \left(Ex_k, (En'_k)^2 \right), m=1, 2, \dots, num \quad (9)$$

Ex_k : expectation (all-0 vector); num : cloud droplet count ($num = \lceil 10 \cdot \exp(t/T) \rceil$); cloud droplets are randomly selected for mutation.

$$X'_{i,k} = X_{i,k} + x_{k,selected} \quad (10)$$

$x_{k,selected}$: random cloud droplet; $X_{i,k}$: original k-dim value of the i -th individual. This strategy boosts panic group diversity.

Optimal attraction strategy

Exploration-to-exploitation transition: populations self-adjust via collective decision-making. Optimal attraction strategy: elite individuals steer the population to accelerate convergence. Individual acceleration is governed by Eq. (11).

$$a_{i,j} = (v_{i,j}^2 - v_{0i,j}^2) / (2w_j) \quad (11)$$

$v_{i,j}$: velocity score (random $\in [lb_j, ub_j]$); $v_{0i,j}$: initial velocity (all-0 matrix); w_j : adaptive Levy weights per dimension (step controller).

Inter-individual attraction coefficients are derived from acceleration and defined as nonlinearly transformed mass-acceleration products (Eq. 12).

$$F_{i,j} = |\lg(M_{i,j} a_{i,j})| \cdot t/T \quad (12)$$

$M_{i,j}$: quality score of the i -th individual in j -th dimension (random $\in [lb_j, ub_j]$). Attraction coefficients are embedded in the update process to amplify elite-to-ordinary attraction for accelerated convergence. Example formula in exploitation phase:

$$x_{i,j}^{new} = x_{i,j} + m_1 \cdot w_1 \cdot (F_{i,j} + E_j - x_{i,j}) + m_2 \cdot w_2 \cdot (x_{rand,j} - x_{i,j}) \quad (13)$$

Parameters in Eq. (13) retain Sec.II.A definitions; updates enhanced via optimal attractiveness strategy for exploration-phase groups (calm, herding, panic).

Empirical modal decomposition with time-varying filtering

TVF-EMD adaptively extracting multi-scale wind speed signals (time-frequency domains). This suppresses noise and resolves mode separation/intermittency issues.

Algorithm steps

TVF-EMD decomposition: three steps:

- (1) local cutoff frequency estimation,
- (2) time-varying filter construction,
- (3) theoretical validation.

- (1) Estimated local cutoff frequency

First utilized: B-spline approximation as low-pass time-varying filter (Eqs. 14, 15).

$$g_m^n(t) = [p_m^n x]_{\downarrow m} b_m^n(t), \quad b_m^n(t) = \beta^n(t/m) \quad (14)$$

$$p_n^m = [([b_n^m b_m^n]_{\downarrow m} b_m^n(t))], \quad b_n^m(t) = \beta^m(t/n) \quad (15)$$

p_n^m : prefilter; $\beta^n(t)$: B-spline (nodes: m ; order: n); $[\cdot]_{\downarrow m}$: m -factor downsampling; $[\cdot]_{\uparrow m}$: m -factor upsampling. Applied to analyzed signal $X(t)$: Hilbert transform³³ extracts amplitude $H(t)$ and instantaneous phase $\varphi(t)$.

$$Z(t) = H(t) \cdot \exp(j\varphi t) = h_1(t) \cdot \exp(j\varphi_1(t)) + h_2(t) \cdot \exp(j\varphi_2(t)) \quad (16)$$

Locate maxima $H(\{t_{\max}\})$ and minima $H(\{t_{\min}\})$; construct interpolation curves ($\beta_1(t)/\beta_2(t)$) to compute instantaneous mean ($h_1(t)$) and envelope ($h_2(t)$).

$$h_1(t) = [\beta_1(t) + \beta_2(t)]/2 \quad (17)$$

$$h_2(t) = [\beta_2(t) - \beta_1(t)]/2 \quad (18)$$

An expression for the local cutoff frequency $\varphi'_{bis}(t)$ is obtained as Eq. (19):

$$\varphi'_{bis} = [\varphi'_1(t) + \varphi'_2(t)]/2 = [\mu_2(t) - \mu_1(t)]/4h_1(t)h_2(t) \quad (19)$$

Finally, the local cutoff frequency φ'_{bis} is readjusted.

- (2). Constructing time-varying filters.

Signal reconstruction is performed by localizing the cutoff frequency φ'_{bis} , as given by Eq. (20).

$$l(t) = \cos \left[\int \varphi'_{bis}(t) dt \right] \quad (20)$$

Construct B-spline with $l(t)$'s extrema (cutoff frequency= φ'_{bis}), filter $X(t)$ to obtain $m(t)$.

- (3). Guidelines for the Cessation of Judgment.

θ evaluates residual signal against stopping criterion (Eq. 21).

$$\theta(t) = B_{loughlin}/\varphi_{avg}(t) \quad (21)$$

If $\theta(t) < \xi$ (bandwidth threshold), residual signal is narrowband (IMFs). $\varphi_{avg}(t)$ (weighted avg. freq.) & $B_{loughlin}$ (threshold) defined by Eqs. (22) and (23).

$$\varphi_{avg}(t) = \frac{h_1^2(t)\varphi'_1(t) + h_2^2(t)\varphi'_2(t)}{h_1^2(t) + h_2^2(t)} \quad (22)$$

$$B_{loughlin}(t) = \sqrt{\frac{h_1^2(t) + h_2^2(t)}{h_1^2(t) + h_2^2(t)} + \frac{h_1^2(t)h_2^2(t)(\varphi'_1(t) - \varphi'_2(t))^2}{(h_1^2(t) + h_2^2(t))^2}} \quad (23)$$

From the above analysis, searching for the optimal parameter combination matching the signal to be analyzed is the key to the TVF-EMD method. This study employs the IESC to optimize the parameters of TVF-EMD.

Fitness function

For wind speed series with high volatility/intermittency, suppressing shock components during decomposition is critical to reduce uncertainty. This study adopts kurtosis minimization as IESC’s fitness function to optimize TVF-EMD, determining optimal B-spline order n and bandwidth threshold ξ . Figure 1 shows the optimization flowchart; kurtosis is defined in Eq. (24):

$$K = \sum_{i=1}^n (x_i - \mu)^4 / n\sigma^4 \tag{24}$$

n : IMFs count; μ : overall modes’ expectation; σ : overall modes’ std. dev.

Regarding the wind speed data, the lower the kurtosis, the better the fitness.

Multi-scale predictive modeling

After the original data undergoes data preprocessing, it is decomposed into multiple intrinsic mode functions (IMFs) via the IESC-optimized TVF-EMD method detailed in section “Empirical modal decomposition with time-varying filtering”, with these IMFs further categorized into high-, medium-, and low-frequency modes. The direct basis for this frequency division lies in the instantaneous frequency characteristics of each IMF: specifically, the instantaneous frequency $\phi'(t)$ of each component is derived through the Hilbert transform, and its statistical distribution across the entire time span determines the corresponding frequency level. Additionally, this classification aligns with the inherent uncertainty and dynamic features of each mode: high-frequency modes exhibit prominent volatility and nonlinearity with the highest uncertainty; medium-frequency modes present moderate complexity and stable oscillatory patterns; while low-frequency modes demonstrate smooth, near-linear trends with strong regularity and minimal uncertainty.

Given the differentiated frequency-domain characteristics and uncertainty levels of the decomposed modes, a one-size-fits-all modeling approach is infeasible—overly complex models tend to overfit low-frequency sequences, while simplistic models fail to capture the intricate dynamics of high-frequency data. To address this challenge, a tailored multi-model prediction strategy is adopted: ELM is selected for low-frequency modes due to its fast training speed and robust performance on steady, near-linear variations; classical LSTM networks are employed for medium-frequency modes, as they effectively balance representational capacity and computational efficiency for moderately complex oscillatory patterns; and XLSTM is utilized for high-frequency modes, leveraging its enhanced memory and attention mechanisms to excel at capturing intricate temporal dependencies in highly volatile and nonlinear data. This mode-specific modeling strategy maximizes the prediction efficiency of individual components and ultimately improves the overall accuracy of wind speed series forecasting.

Extreme learning machine (ELM)

ELM is a single-hidden-layer feedforward neural network, which boasts fast training speed and strong stability. These inherent advantages make it suitable for low-frequency wind speed data post-decomposition.

As a single-hidden-layer feedforward neural network, ELM.

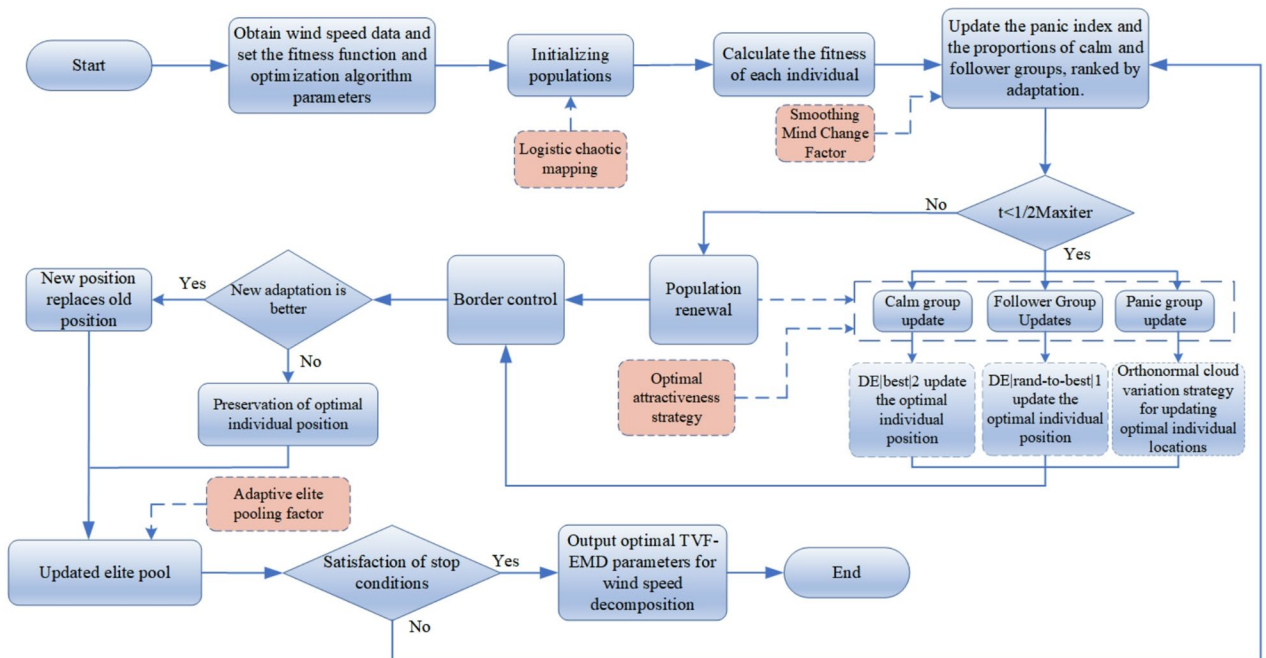


Fig. 1. Flowchart of TVF-EMD optimization by IESC.

Hidden layer: Compute neuron outputs using Sigmoid as the activation function with random weights and biases.

$$h_i = g(w_i x + b_i) \tag{25}$$

Output layer: Solve weights β via least squares, then calculate final output linearly.

$$\beta = H^\dagger Y \tag{26}$$

Long short-term memory (LSTM) networks

With a simpler structure, LSTM enabling robust capture of temporal dependencies in time series data. This makes it particularly effective for medium-frequency signal prediction, so it is applied here to forecast mid-frequency wind speed components. Its unit contains three gates: input gate, forget gate, and output gate. Calculations per time step follow:

$$f_t = \sigma(W_f h_{t-1} + W_f x_t + b_f) \tag{27}$$

f_t is the forget gate output; W_f/b_f denote its weights/bias. σ is the Sigmoid function. h_{t-1} is previous hidden state; x_t is input sequence.

Other gates (input/output) and memory cell state follow analogous logic: input gate filters input, output gate generates hidden state, and memory cell C_t updates via forget/input gate outputs.

Expanding long and short-term memory networks

XLSTM enhances LSTM with two variants: scalar (sLSTM) and matrix (mLSTM). For high-frequency wind speed data, which is prone to strong randomness and uncertainty, sLSTM is selected for this study due to its superior ability to capture rapid temporal fluctuations.

The stabilized input and forgetting gates are calculated as follows:

$$i'_t = \exp(\log(i_t) - m_t) \tag{28}$$

$$f'_t = \exp(\log(f_t) + m_{t-1} - m_t) \tag{29}$$

The stabilized input/forget gate outputs prevent exponential activation's numerical explosions, keeping values stable.

As Fig. 2 illustrates, This study constructs a multi-scale wind speed prediction model via frequency-specific predictors. Wind speed data are decomposed using an IESC-optimized TVF-EMD algorithm, partitioned into high/medium/low-frequency bands. Meteorological data are normalized with decomposed series. High-frequency data feed XLSTM; medium to LSTM; low to ELM. Final prediction integrates all band outputs.

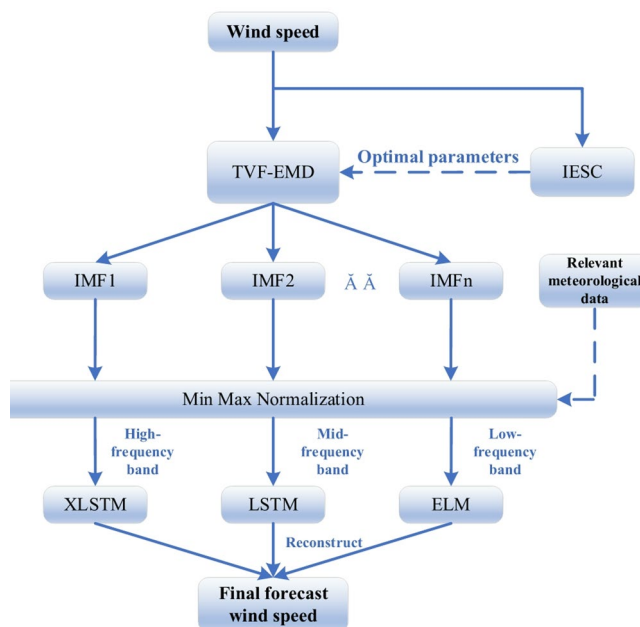


Fig. 2. Multi-scale wind speed prediction mode.

Model evaluation indicators IESC-optimized TVF-EMD model

For TVF-EMD, this study evaluates model performance through mode count, decomposition time, and IC (reconstruction error index). Higher mode counts accumulate prediction errors, raising total error; excessively low modes cause inadequate trend-noise separation, hindering pattern capture. Optimal mode count balances noise suppression and error minimization via effective decomposition. IC magnitude quantifies decomposition deviation (Eq. 30).

$$IC = \sqrt{\sum_{t=0}^T \left[x(t) - \sum_{i=1}^N c_i(t) \right]^2 / T} \quad (30)$$

t denotes the t -th eigenmode index, $x(t)$ (wind speed sequence) and $c_i(t)$ (eigenmode function). T : sample count; N : mode count. Decomposition accuracy inversely correlates with IC—lower IC signifies superior decomposition. When TVF-EMD's N and IC reach joint optimality, shorter decomposition time implies higher efficiency and enhanced model performance.

Multi-scale predictive model

Wind speed prediction accuracy is evaluated via prediction-actual fit and error metrics. Higher R^2 (Coefficient of Determination) and EV (Explained Variance) indicate stronger model fit, while lower MAE (Mean Absolute Error), MSE (Mean Squared Error), and RMSE (Root Mean Squared Error) signify reduced errors. The metrics are calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (31)$$

$$EV = 1 - Var(y_i - \hat{y}_i) / Var(y_i) \quad (32)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (33)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad RMSE = \sqrt{MSE} \quad (34)$$

Where y_i is the actual value, \hat{y}_i represents the predicted value, \bar{y} is the mean value, n is the number of samples, and $Var(\cdot)$ indicates the variance calculation.

Calculus analysis

This section assesses IESC-TVF-EMD decomposition performance and conducts multi-scale framework-based simulation analysis. The evaluation covers two components:

- (1) Data processing.
- (2) Experimental results and discussion.

Data processing

This study employs 2019 wind speed data from a Xinjiang wind farm³⁵ with 15-minute resolution. January dataset is selected with 2976 data points, and 70 m hub-height wind speed serves as the prediction target.

Data is split into 70% training, 10% validation (70th–80th percentiles), and 20% test sets. Input features include 10 m, 30 m, 50 m, 70 m hub-height wind speed and 10 m, 70 m wind direction data.

The data have been made open-source and deposited on the Zenodo platform, it can be obtained from: <https://doi.org/10.5281/zenodo.16535301>.

All data generated or analysed during this study are included in supplementary information files.

Experimental results and discussion

This subsection benchmarks TVF-EMD optimized by competing algorithms against the proposed IESC-enhanced variant, simulates and analyzes wind speed modal predictions across high, medium, low frequency bands, and validates the multi-scale framework via multi-step forecasting.

Experimental analysis of IESC-optimized TVF-EMD decomposition

To determine the optimal objective function for IESC-optimized TVF-EMD decomposition, we conducted comparative experiments using six candidate objective functions (Envelope spectral peak factor, Kurtosis, Energy entropy, etc.), with the results summarized in Table 1.

This experiment employs IESC to optimize TVF-EMD, with ESC, Gray Wolf Algorithm (GWO), Particle Swarm Optimization (PSO), and Differential Evolution (DE) benchmarked for fitness values and decomposition performance. All algorithms are assessed under six fitness functions: envelope spectral peak factor, kurtosis, energy entropy, envelope kurtosis factor, envelope entropy, and average envelope spectral entropy minimization.

The optimization algorithm uses a population size of 30 and 10 iterations, with bandwidth thresholds 0.1–0.8 and B-spline order limits 6–30. January wind speed data (2976 points) are analyzed.

Objective function	Algorithm	Fitness	Num	Time (s)	IC	Threshold	B-spline
Envelope spectral peak factor	IESC	0.010411	8	1.056	3.4976e-16	0.63241	28
	ESC	0.043959	9	1.318	3.9086e-16	0.53591	21
	GWO	0.056615	9	0.9608	3.8822e-16	0.52344	21
	PSO	0.056615	9	0.870	3.9086e-16	0.53307	21
	DE	0.047735	9	0.370	3.7978e-16	0.63686	9
Kurtosis	IESC	1.308	6	0.449	2.0925e-16	0.7583	23
	ESC	1.3952	6	0.954	2.6373e-16	0.76718	10
	GWO	1.4725	6	1.207	2.2055e-16	0.71877	20
	PSO	1.33	9	0.875	4.1303e-16	0.52899	21
	DE	1.4283	10	1.105	5.5867e-16	0.48012	20
Energy entropy	IESC	0.41061	2	0.164	1.5589e-17	0.8	14
	ESC	0.41061	2	0.181	1.5589e-17	0.8	14
	GWO	0.41061	2	0.202	1.5589e-17	0.8	14
	PSO	0.41061	2	0.174	1.5589e-17	0.8	14
	DE	0.41061	2	0.149	1.5589e-17	0.8	14
Envelope kurtosis factor	IESC	5.9241e-05	10	0.484	5.1373e-16	0.48496	10
	ESC	0.00014087	10	1.370	5.1639e-16	0.45253	21
	GWO	0.00013257	10	1.100	6.8001e-16	0.40223	16
	PSO	0.0003623	10	0.648	5.0014e-16	0.45204	14
	DE	0.0079571	10	0.471	4.6952e-16	0.5046	9
Envelope entropy	IESC	6.4878	10	0.839	4.3683e-16	0.52445	7
	ESC	6.8686	10	0.907	5.7185e-16	0.42976	15
	GWO	6.8686	10	0.435	4.3683e-16	0.5265	7
	PSO	7.1013	10	1.985	5.2149e-16	0.49265	21
	DE	7.1013	10	1.103	5.2149e-16	0.49009	21
Average envelope spectral entropy	IESC	7.6083	6	0.230	7.1765e-16	0.57042	10
	ESC	7.6326	7	0.312	8.2006e-16	0.72788	15
	GWO	7.6326	6	0.232	7.1765e-16	0.53014	10
	PSO	7.6356	8	0.722	7.1765e-16	0.47776	18
	DE	7.6326	6	0.249	7.1765e-16	0.54186	10

Table 1. Optimizing decomposition results.

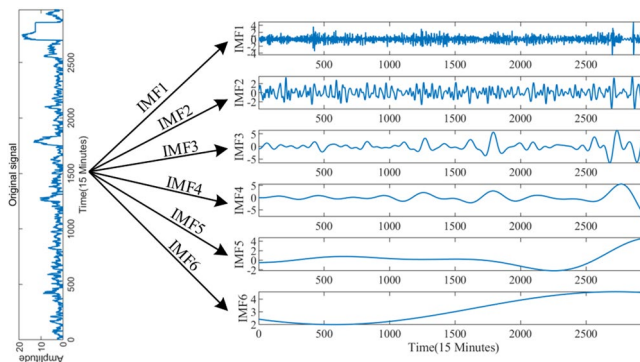


Fig. 3. IESC-optimized TVF-EMD decomposition of modal diagrams.

Among the five algorithms, IESC achieves the lowest fitness values across kurtosis objective functions. Meanwhile, its runtime is notably shorter than that of ESC, GWO, etc. IESC also yields reasonable IMF counts and parameters aligned with TVF-EMD theory. As Table 1 shows, IESC achieves superior fitness across all functions.

As the objective function, kurtosis effectively separates wind speed modes (Sec.II.B). Figure 3 displays the IESC-optimized TVF-EMD modes with faster decomposition than counterparts. Figure 4 visualizes fitness function optimization results, confirming IESC’s effectiveness. Figure 5 plots fitness curves, revealing IESC’s superior initial fitness and convergence across functions.

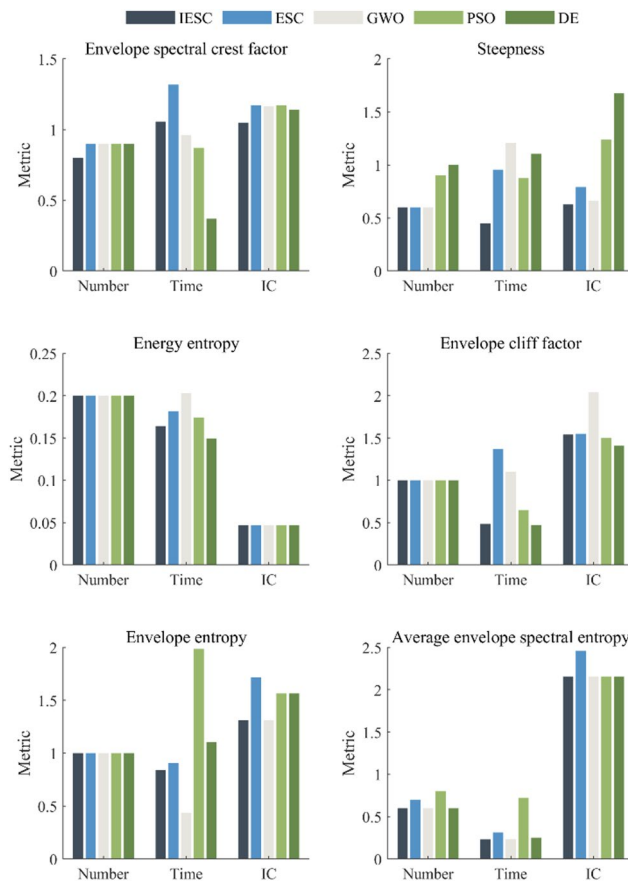


Fig. 4. Optimization results for each fitness function.

Parameter sensitivity analysis

The optimization performance of the IESC algorithm is directly affected by its core parameters, namely population size and number of iterations. In this section, the control variable method is used, where population size (size pop) is set to 20, 30, 40, 50, and 60, and iterations is set to 5, 10, 15, 20, and 25. The experimental data are shown in Table 2.

The results indicate that the fitness value generally decreases with increasing population size and iterations, but the runtime increases significantly. For instance, at sizepop = 50 and Maxiter = 5, the fitness value reaches the minimum of 7.6083, while the runtime is 536.24 s. In contrast, at sizepop = 60 and Maxiter = 25, the runtime is as high as 2194.45 s with a fitness value of 7.6326. Considering both optimization accuracy and computational efficiency, we selected sizepop = 30 and Maxiter = 10 as the default parameters, which yield a fitness value of 7.5861 and a runtime of 517.77 s, representing a good trade-off. The heat map results are shown in Fig. 6.

Cross-dataset stability verification

We selected wind speed data from a Xinjiang wind farm for June 2019 and September 2019 and compared these datasets with the data from January to confirm the stability and repeatability of the IESC-based optimization. The results are showed in Table 3.

The consistent mean values and low standard deviations across datasets indicate that IESC-based optimization is stable and reliable for different wind speed datasets with similar statistical characteristics.

Comparison of computational complexity for large datasets

The large datasets used in the experiment were collected from a 2019 wind speed dataset from the Xinjiang wind farm³⁵, which has a 15-minute resolution and a total of 35,040 data points. Meanwhile, we set two dataset scales with 6-month and 12-month. Considering computational efficiency, all algorithms adopt consistent parameter settings: population size = 30 and Maxiter = 10 with the same optimization objective. Each experiment is repeated 6 times, and the average running time is taken as the final result.

We chose the kurtosis fitness function (Sec.II.B) for comparing the computational complexity of different algorithms when dealing with large scale data.

In Figs. 7 and 8, we compare the convergence of IESC, ESC, DE, GWO, and PSO under 6-month and 12-month datasets. In the 6-month dataset, IESC converges the fastest, stabilizing at 1.349 after 4 iterations and achieving the lowest fitness, while DE fails to converge effectively; In the 12-month dataset, all algorithms show an initial increase in fitness, but IESC still stabilizes quickly, achieving this in 5 iterations and obtaining the

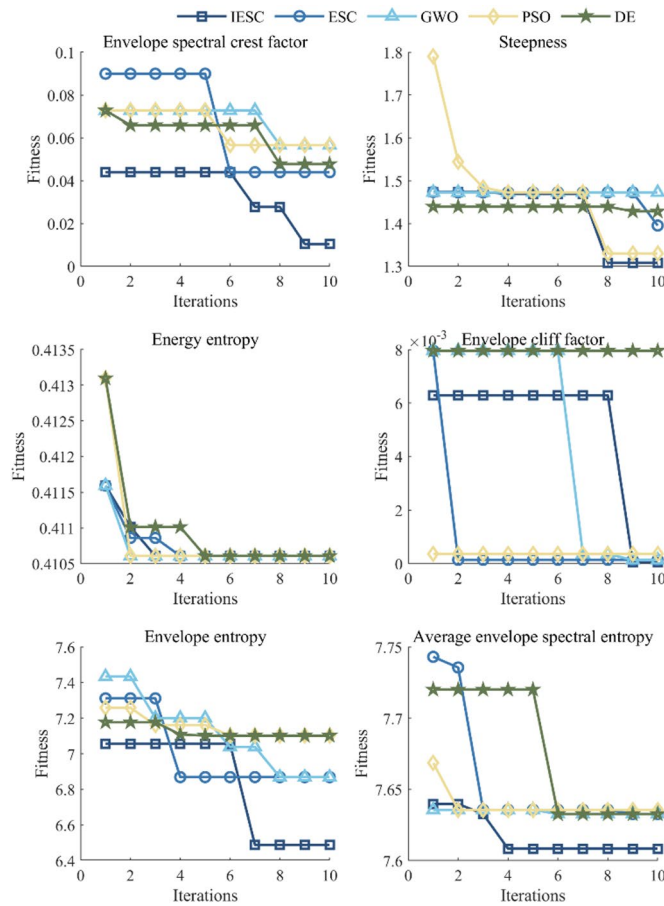


Fig. 5. Fitness curve.

Iterations	Size pop				
	20	30	40	50	60
5	7.7210	7.6407	7.6326	7.6083	7.6020
10	7.6598	7.6272	7.5861	7.6326	7.6170
15	7.6326	7.5861	7.6020	7.6020	7.6326
20	7.6326	7.5923	7.6083	7.6020	7.6020
25	7.7297	7.6020	7.6326	7.5861	7.6326

Table 2. Effects of different population numbers and iterations on the fitness function.

lowest fitness. Its fitness decline is greater than that of ESC. From the 6-month to the 12-month dataset, as the data scale doubles, IESC demonstrates stronger robustness.

Overall, IESC outperforms the other algorithms in terms of convergence speed, precision, and stability across both datasets.

The comparison results of the running time of each algorithm under large datasets are shown in Fig. 9.

In large-scale wind speed prediction scenarios, the computational complexity of the proposed IESC algorithm is lower than that of the original ESC, and its running time is within an acceptable range. Compared with classical algorithms such as DE, GWO, and PSO, IESC sacrifices a small amount of computational efficiency to obtain higher optimization accuracy, which is more in line with the actual demand of wind speed prediction.

Experimental analysis of prediction for each mode

This section evaluates the fit between decomposed wind speed modes and predictors by testing six models (XLSTM, LSTM-Transformer (T-LSTM), Transformer, LSTM, BP neural network, ELM) across six frequency bands, ordered from simple to complex. As Fig. 10 shows, XLSTM outperforms others in high-frequency modes 1–2 with higher EV scores and 1–10% error reduction. Conversely, ELM-being simpler-achieves the lowest accuracy in high/medium bands but excels in low-frequency bands. Complex models show larger deviations,

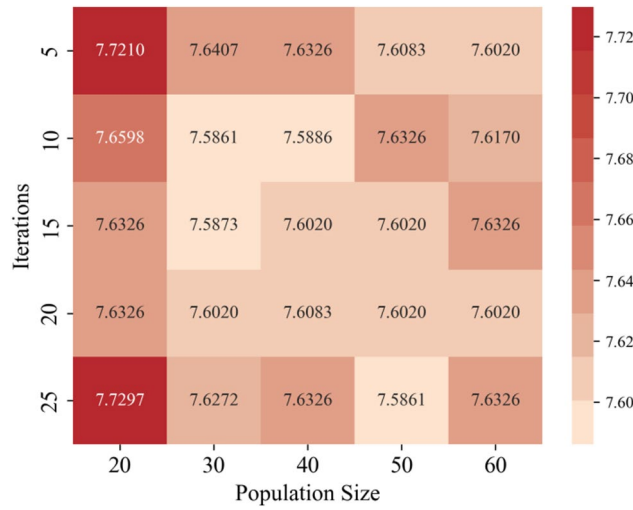


Fig. 6. Heatmap of fitness values under different population sizes and iterations for the IESC algorithm.

Dataset	B-spline	Threshold	Fitness	IC
January	23	0.7583	1.3080	2.09e-16
June	23	0.7496	1.3120	2.15e-16
September	23	0.7617	1.2980	2.11e-16

Table 3. IESC-optimized TVF-EMD: parameters & performance across datasets.

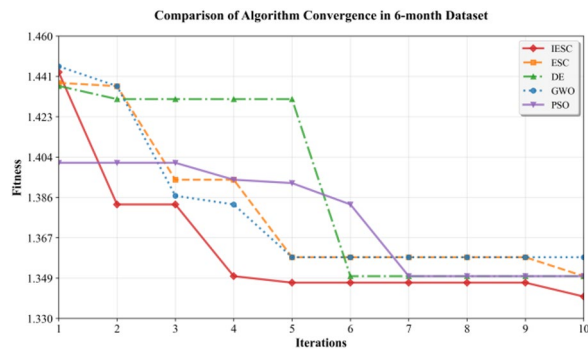


Fig. 7. Comparison of algorithm convergence in 6-month datasets.

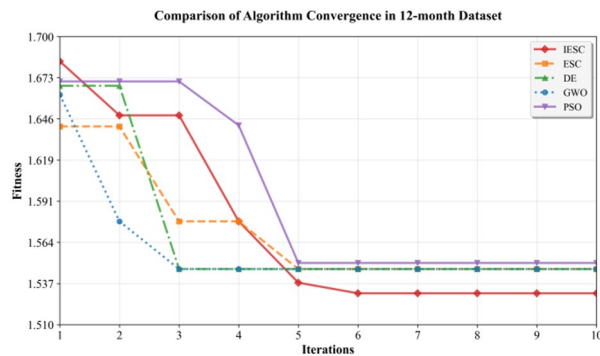


Fig. 8. Comparison of algorithm convergence in 12-month datasets.

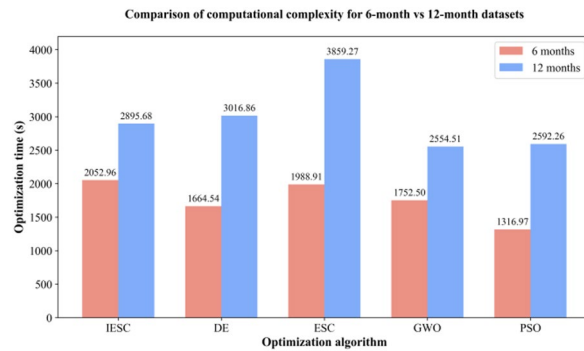


Fig. 9. Comparison of computational complexity for 6-month vs. 12-month datasets.

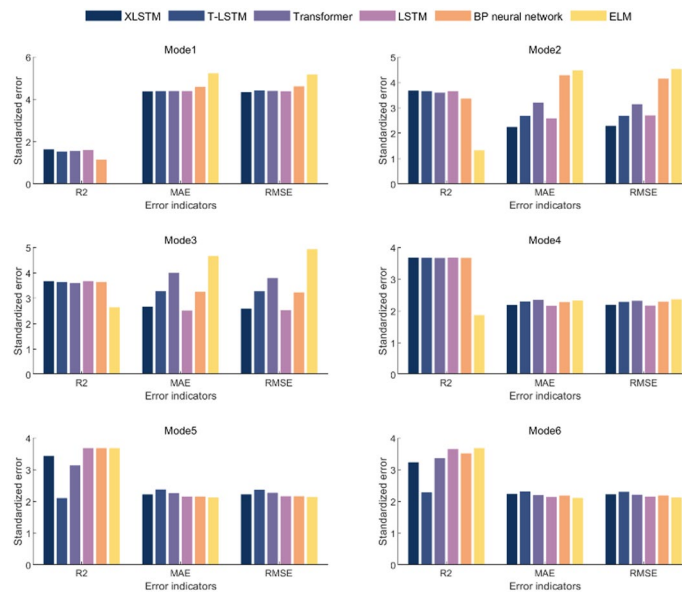


Fig. 10. Predictive metrics for each model.

while ELM’s predictions align closely with actual data. For medium-frequency modes III-IV, LSTM delivers superior performance.

XLSTM and T-LSTM overfit due to high complexity, while ELM underfits. Thus, LSTM is preferred for medium-frequency bands, as Fig. 11 demonstrates.

Experimental analysis of over-the-top multi-step prediction

The prediction performance of the proposed multi-scale model was evaluated by benchmarking existing XLSTM, T-LSTM, Transformer, LSTM, BP neural network, and ELM models across 1-15-step-ahead forecasts at selected intervals (1,3,6,9,12,15), with results summarized in Table 4.

In contrast, the proposed multi-scale model is observed to exhibit only a 0.21 score decrease, indicating superior fit maintenance in multi-step predictions, as is further validated in Fig. 12. Regarding prediction error, the XLSTM model is characterized by MAE and RMSE increases of 2.24 and 2.82 respectively, while the multi-scale model is limited to 1.42 and 1.78 increases. These values are nearly halved compared to XLSTM, confirming the multi-scale model’s enhanced capability for long-term predictions through lower error indices, as demonstrated in Fig. 13.

As shown in Table 4, the proposed multi-scale model exhibits marginally superior performance in 1-step prediction, with MAE and RMSE reductions of 0.9 and 1.25 respectively compared to counterparts, demonstrating significant wind speed error reduction. With increasing prediction steps and prediction accuracy deteriorate substantially. For 15-step predictions, the top-performing XLSTM model shows a 0.5 score decline, indicating severe model fit degradation.

Performance beyond 15-step prediction: bias drift, error accumulation, and stability analysis

Regarding the performance beyond 15-step prediction, the extended results (20/25/30 steps) show no evidence of continuous error accumulation or significant bias drift, as demonstrated in Fig. 14. Specifically, R² values

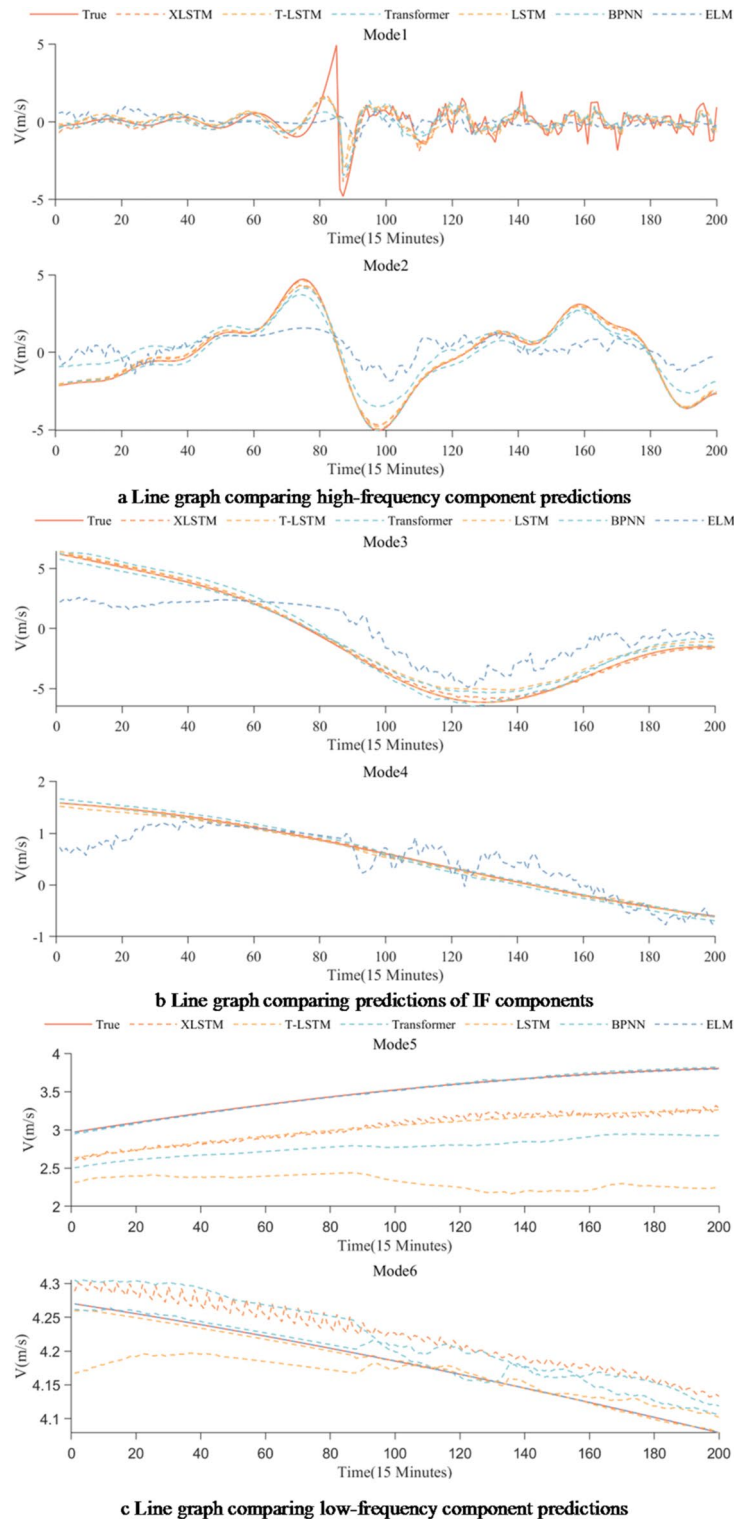


Fig. 11. Line graph comparing the prediction of each mode.

are -17.0677 , -16.1272 , and -16.8185 for 20/25/30 steps, respectively, improving from -21.2226 at 15 steps without sustained degradation. MAE decreases from 5.6119 (15 steps) to 5.0019 (25 steps) and slightly rises to 5.2492 (30 steps), indicating non-cumulative error behavior. No consistent upward/downward trend in MAE or R^2 confirms no systematic bias drift. The relatively stable fluctuations of key metrics demonstrate the model maintains acceptable long-term stability beyond 15 steps. The suboptimal negative R^2 may stem from unmodeled nonlinear wind speed characteristics in extended horizons, while non-cumulative errors validate the

Pacemaker	Algorithms	R2	MAE	RMSE
One-step	Multi-scale	0.9795	0.5109	0.7541
	XLSTM	0.9613	0.7298	1.0359
	T-LSTM	0.8785	1.2493	1.8364
	Transformer	0.8049	1.6201	2.3277
	LSTM	0.9260	0.9592	1.4333
	BPNN	0.8931	1.2723	1.7231
	ELM	0.8068	1.6547	2.3162
Three-step	Multi-scale	0.9716	0.6295	0.8868
	XLSTM	0.9076	1.1061	1.6000
	T-LSTM	0.5817	2.4275	3.4055
	Transformer	0.5101	2.7783	3.6851
	LSTM	0.8139	1.7814	2.2713
	BPNN	0.7652	1.9385	2.5513
	ELM	0.6608	2.2478	3.0663
Six-step	Multi-scale	0.9504	0.8778	1.1705
	XLSTM	0.8199	1.5862	2.2321
	T-LSTM	0.7535	1.9369	2.6112
	Transformer	0.4943	2.7614	3.7404
	LSTM	0.6709	2.2734	3.0175
	BPNN	0.6326	2.4751	3.1881
	ELM	0.5601	2.5949	3.4887
Nine-step	Multi-scale	0.8592	1.4923	1.9715
	XLSTM	0.6451	2.4143	3.1309
	T-LSTM	0.6169	2.4977	3.2529
	Transformer	0.1656	3.6281	4.8008
	LSTM	0.5162	2.8125	3.6555
	BPNN	0.4299	2.9634	3.9683
	ELM	0.3084	3.2148	4.3707
Twelve-steps	Multi-scale	0.8266	1.6558	2.1899
	XLSTM	0.5394	2.7137	3.5688
	T-LSTM	0.4656	2.8243	3.8441
	Transformer	0.4505	3.0939	3.8980
	LSTM	0.4375	2.9399	3.9440
	BPNN	0.5188	2.8075	3.6479
	ELM	0.3439	3.1376	4.2594
Fifteen-steps	Multi-scale	0.7685	1.9365	2.5327
	XLSTM	0.4640	2.9637	3.8545
	T-LSTM	0.2175	3.5387	4.6570
	Transformer	0.1953	3.6283	4.7226
	LSTM	0.3798	3.2676	4.1462
	BPNN	0.4945	2.7637	3.7430
	ELM	0.2527	3.3090	4.5511

Table 4. Prediction results for each model.

IESC-optimized TVF-EMD and multi-scale framework's effectiveness. predictions, the top-performing XLSTM model shows a 0.5 score decline, indicating severe model fit degradation.

Conclusion

This study proposes a multi-scale wind speed prediction model based on TVF-EMD optimized by IESC.

To address slow convergence and local optima in heuristic algorithms like GA, PSO and GWO, IESC integrates four key improvements—logistic chaotic initialization, elite pool grouping, differentiated mutation and optimal attraction. Under kurtosis minimization, its fitness value reaches 1.308, 6.2%, 11.3% and 8.4% lower than ESC, GWO and DE respectively; decomposition time is only 0.449s, 53% faster than ESC; reconstruction error IC is $2.0925e-16$, enabling precise optimization of TVF-EMD parameters including B-spline order and bandwidth threshold.

Based on the decomposition of different IMFs, the XLSTM-LSTM-ELM frequency-specific strategy is proposed: XLSTM captures high-frequency noise with MAE reduced by 10%-15%, LSTM adapts to medium-

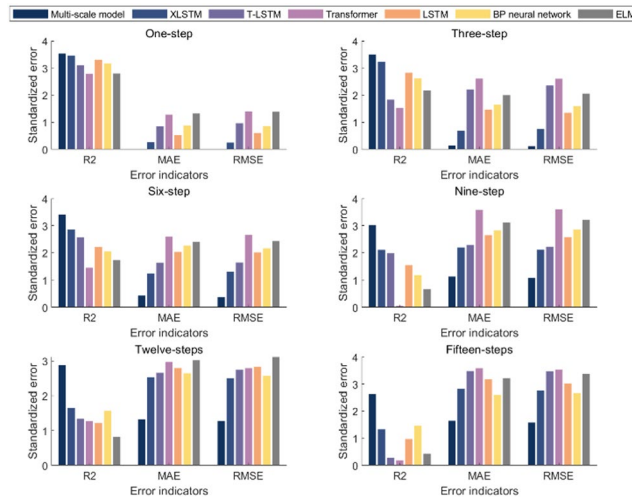


Fig. 12. Prediction results for different step sizes.

frequency oscillations with R^2 of 0.8139, and ELM fits low-frequency trends, advancing the refined paradigm of multi-scale prediction.

In addition, the model shows excellent multi-step prediction performance: It achieves a 29.8% lower 1-step MAE (0.5109) and a 65.6% higher 15-step R^2 (0.7685) compared to XLSTM. More importantly, its error growth (MAE and RMSE) is significantly restrained at only 63.5% and 63.1% of XLSTM's rates. This results in a controlled 1–15 step error growth of under 12% and a 35% slower R^2 degradation. The model also maintains excellent long-term stability over extended horizons without continuous error accumulation.

The model's 3.75-hour-ahead forecasting capability improving TSO-DSO coordination efficiency, reducing MGs power imbalance and mitigating wind power volatility's impact on grid dispatching, facilitating the integration of higher shares of renewable energy. Building on this, future work will develop dynamic modal division using real-time volatility and incorporate multi-source meteorological data such as air pressure and temperature with spatial correlations, with the goal of extending the model to offshore wind farms to address their higher volatility.

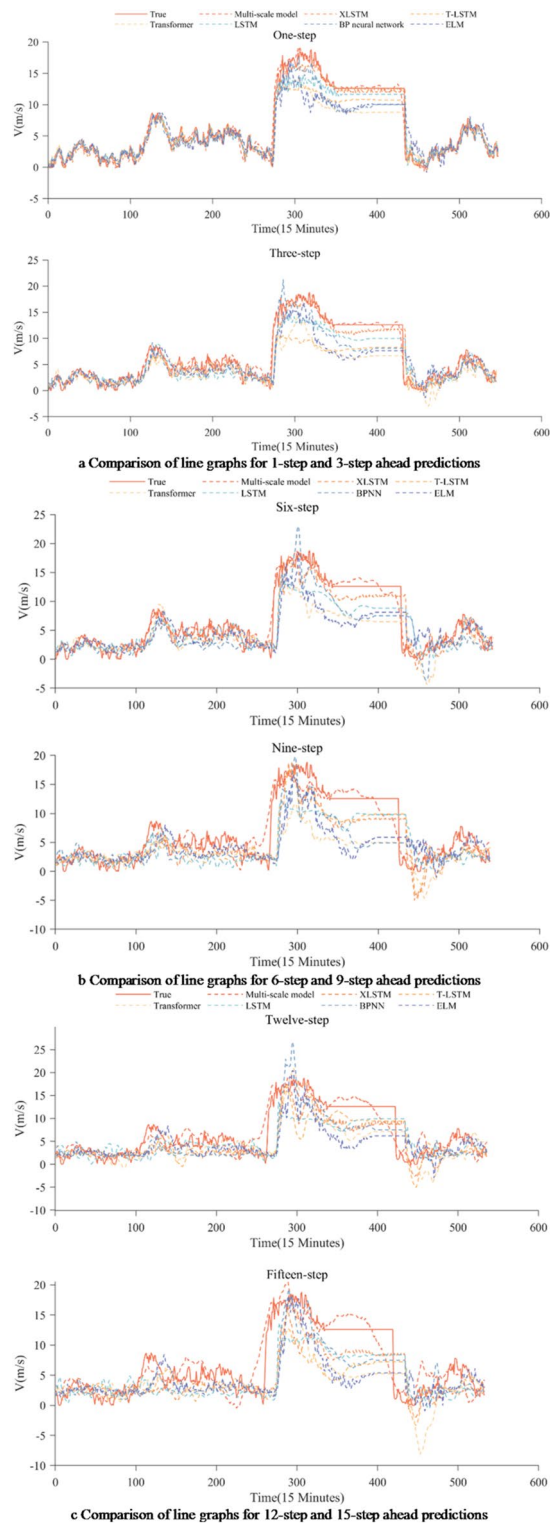


Fig. 13. Comparison of line graphs of over-the-horizon multi-step predictions.

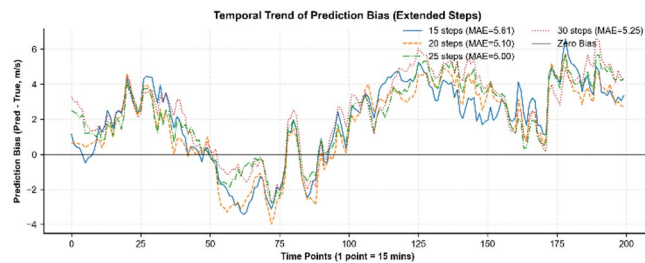


Fig. 14. Beyond 15-step prediction.

Data availability

The data have been made open-source and deposited on the Zenodo platform, it can be obtained from: <https://doi.org/10.5281/zenodo.16535301>.

Received: 21 July 2025; Accepted: 6 January 2026

Published online: 09 January 2026

References

- Mansouri, S. A., Paredes, Á., González, J. M. & Aguado, J. A. A three-layer game theoretic-based strategy for optimal scheduling of microgrids by leveraging a dynamic demand response program designer to unlock the potential of smart buildings and electric vehicle fleets. *Appl. Energy*. **347**, 0306–2619 (2023).
- Zhou, X., Mansouri, S. A., Jordehi, A. R., Tostado-Véliz, M. & Jurado, F. A three-stage mechanism for flexibility-oriented energy management of renewable-based community microgrids with high penetration of smart homes and electric vehicles. *Sustainable Cities Soc.* **99**, 2210–6707 (2023).
- Mansouri, S. A. et al. An interval-based nested optimization framework for deriving flexibility from smart buildings and electric vehicle fleets in the TSO-DSO coordination. *Appl. Energy*. **341**, 0306–2619 (2023).
- Le, X. et al. Optimization study of grid access for wind power system considering energy storage. *Scalable Comput.: Pract. Exp.* **25** (2), 721–728 (2024).
- Tyass, I. et al. Wind speed prediction based on seasonal ARIMA model. In *E3S Web of Conferences* 336: 00034 (EDP Sciences, 2022).
- Lenora, K. M., Abeyesundara, S. P. & Perera, K. A hybrid model for wind speed prediction in Anuradhapura. *Sri Lanka* (2021).
- Buturache, A. N. & Stancu, S. Wind energy prediction using machine learning. *Low Carbon Econ.* **12** (01), 1 (2021).
- Uzair, M., Shah, I. & Ali, S. An adaptive strategy for wind speed forecasting under functional data horizon: a way towards enhancing clean energy. *IEEE Access*. (2024).
- Li, Z. et al. Short-term wind power prediction based on extreme learning machine with error correction. *Prot. Control Mod. Power Syst.* **1** (1), 1–8 (2016).
- Hesty, N. W. et al. Promoting wind energy by robust wind speed forecasting using machine learning algorithms optimization (2024).
- Al-Selwi, S. M. et al. LSTM inefficiency in long-term dependencies regression problems. *J. Adv. Res. Appl. Sci. Eng. Technol.* **30** (3), 16–31 (2023).
- Vazhayil, A. & DeepProteomics, K. P. S. Protein family classification using shallow and deep networks. arxiv preprint arxiv:1809.04461 (2018).
- Ahmed, N. Y. et al. An efficient deep learning approach for DNA-binding proteins classification from primary sequences. *Int. J. Comput. Intell. Syst.* **17** (1), 88 (2024).
- Canducci, M. et al. Tracking the temporal-evolution of supernova bubbles in numerical simulations. In *Intelligent Data Engineering and Automated Learning-IDEAL 2021: 22nd International Conference, IDEAL 2021, Manchester, UK, November 25–27, 2021, Proceedings* 22 493–501 (Springer International Publishing, 2021).
- Zhu, T., Wang, W. & Yu, M. Short-term wind speed prediction based on FEEMD-PE-SSA-BP. *Environ. Sci. Pollut. Res.* **29** (52), 79288–79305 (2022).
- Huang, Y. et al. Multi-step wind speed forecasting based on ensemble empirical mode decomposition, long short term memory network and error correction strategy. *Energies* **12** (10), 1822 (2019).
- Sharma, J., Faiyaz, H. & Manhar, A. *Short Term Wind Forecasting Using Machine Learning Models with Noise Assisted Data Processing Method* (springer, 2023).
- Lv, S., Wang, L. & Wang, S. A hybrid neural network model for short-term wind speed forecasting. *Energies* **16** (4), 1841 (2023).
- Li, H., Li, Z. & Mo, W. A time varying filter approach for empirical mode decomposition. *Sig. Process.* **138**, 146–158 (2017).
- Ye, X. et al. An adaptive optimized TVF-EMD based on a sparsity-impact measure index for bearing incipient fault diagnosis. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2020).
- Yang, D. et al. An improved genetic algorithm and its application in neural network adversarial attack. *Plos One*. **17** (5), e0267970 (2022).
- Sekyere, Y. O. M., Effah, F. B. & Okyere, P. Y. An enhanced particle swarm optimization algorithm via adaptive dynamic inertia weight and acceleration coefficients. *J. Electron. Electr. Eng.* **2024**, 53–67 (2024).
- Hafeez, G. et al. An innovative optimization strategy for efficient energy management with day-ahead demand response signal and energy consumption forecasting in smart grid using artificial neural network. *IEEE Access*. **8**, 84415–84433 (2020).
- Aala Kalananda, V. K. R. & Komanapalli, V. L. N. A competitive learning-based grey Wolf optimizer for engineering problems and its application to multi-layer perceptron training. *Multimedia Tools Appl.* **82** (26), 40209–40267 (2023).
- Ouyang, K. et al. Escape: an optimization method based on crowd evacuation behaviors. *Artif. Intell. Rev.* **58** (1), 19 (2024).
- Shibayama, H. & Nakamura, S. Complication appraised by wave form of environmental noise. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering* 2877–2880 (1996).
- Kushwah, V., Wadhvani, R. & Kushwah, A. K. Trend-based time series data clustering for wind speed forecasting. *Wind Eng.* **45** (4), 992–1001 (2021).
- Beck, M. et al. Xlstm: extended long short-term memory (2024). arxiv preprint arxiv:2405.04517.

29. Rani, R. H. J. & Victoire, T. A. A. Training radial basis function networks for wind speed prediction using PSO enhanced differential search optimizer. *PLoS One*. **13** (5), e0196871 (2018).
30. Zhao, Y. et al. Apple grading based on IGWO optimized support vector machine. In *Proceedings of International Conference on Artificial Life and Robotics* (2023).
31. Deng, L. et al. Dual mutations collaboration mechanism with elites guiding and inferiors eliminating techniques for differential evolution. *Soft. Comput.* **2022**, 1–18 (2022).
32. Reali F, Priami C, Marchetti L. Optimization algorithms for computational systems biology[J]. *Frontiers in Applied Mathematics and Statistics*, **2017**, 3: 6 (2017).
33. Chen, M. Automatic console image processing aided by improved particle swarm computing intelligent algorithm. *Math. Probl. Eng.* **2022** (1), 3475806 (2022).
34. Huang, N. E. 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.* **454** (1971), 903–995 (1998).
35. Haili, Z. et al. Multi-scale wind speed prediction model based on improved escape algorithm for optimizing time-varying filtering empirical modal decomposition [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.16535301> (2025).

Author contributions

Haili Zheng and Qian Wu wrote the main manuscript text and provided experimental analysis, they contributed equally to this work. Xuanhan Lv prepared Figs. 1, 2 and 3. Peitong Zeng prepared Figs. 4, 5, 6, 7 and 8. Xiaoli Zhu prepared all tables. Yupan Zeng provided suggestions and improvement measures. All authors reviewed the manuscript.

Funding

This work is supported by program for scientific research start-up funds of Guangdong Ocean University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-35505-6>.

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026