ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 3 - 2026

Ultra-Short-Term Wind Power Forecasting Based on Variational Mode Decomposition and XGBoost Ensemble

Hui Yuan[1,*], *Minjie Chai*[2], *Siqing Xu*[1], *Jinsong Li*[1] *and Jinwan Zheng*[1]

[1] Electric Power Research Institute, State Grid Shanxi Electric Power Co., Ltd., Taiyuan, 030001, Shanxi, China
[2] Jincheng Power Supply Branch, State Grid Shanxi Electric Power Co., Ltd., Jincheng, 048000, Shanxi, China

SUMMARY: *A VMD-XGBoost ensemble method for enhancing the accuracy of ultra-short-term wind power forecasting at 2-4 hour horizons. The original power series is decomposed into intrinsic mode functions by variational mode decomposition (VMD), and separate XGBoost models, with lagged features, are built for each component, with final forecasts obtained by summation. Using 3.3 years of 15-min wind farm data, the proposed method achieves an R2 of 0.8922 and an RMSE of 21.815 MW at the 4-h horizon, outperforming raw XGBoost by 18.1% in R2 and 33.6% in RMSE. Inference time remains below 0.05 s, confirming real-time applicability.*

KEYWORDS: *wind power forecasting; variational mode decomposition; XGBoost, ensemble learning; ultra-short-term forecasting*

1 Introduction

As the "dual carbon" goals are being vigorously pursued, wind power's share in the power system continues to grow [1]. However, the inherent intermittency and volatility of wind power pose significant challenges to grid dispatch[2–4]. Ultra-short-term wind power forecasting (0-4 h) provides key technical support for real-time dispatch, rolling optimization, and reserve capacity allocation. Higher forecasting accuracy directly influences dispatch costs and curtailment rates [5–7].

In ultra-short-term forecasting, the baseline performance of the persistence model at the one-minute level is very high because of the continuity of the evolution of the wind field and the inertia of the wind turbines. It serves as a litmus test for the efficacy of arbitrary models. The autocorrelation nature of the power series deteriorates over 1 h forecast horizon and the driving role of local features of meteorological observations like wind speed and wind direction is significantly enhanced. It is no longer adequate to merely extrapolate pure inertia to meet accuracy requirements. As a result, the use of machine learning models (XGBoost, random forest, etc.) has been studied extensively to capture non-linear meteorological-to-power mapping relationships [8–10].

XGBoost is a great option for wind power forecasting because of its fast training, having few hyperparameters to tune, and estimating feature importance natively [11]. Nonetheless, the original power series shows strong non-stationarity and multi-scale characteristics, making it difficult to fully exploit the inherent patterns using the raw series as input for XGBoost. Signal

*yuanhuidky@163.com
https://doi.org/10.65102/is20261299
Published: 8 July, 2026

decomposition schemes, such as the empirical mode decomposition (EMD) and the variational mode decomposition (VMD), can greatly reduce the modeling difficulty by decomposing the original series into several relatively stationary intrinsic mode functions (IMFs) [12-14]. Among these methods, the VMD adaptively decomposes the signal into a specified number of finite-bandwidth text-time-frequency IMFs by constructing and solving a variational problem, effectively avoiding mode mixing [15, 16].

Yet, research on coupling VMD with XGBoost for ultra-short-term wind power forecasting is scarce. There is a lack of systematic validation across multiple forecasting horizons (15 min to 4 h) and over a long timeframe (over 3 years). In addition, the majority of existing literature attempts to design various ensemble or deep learning architectures without organizing the decomposition-based feature enhancement [17], and the resulting studies do not clearly expose how the modal contributions evolved forecasts horizon after VMD decomposition. Another line of work employs a spatiotemporal hypergraph learning approach [18], which achieves high expressiveness but has considerable model complexity and computational burden, which obstructs engineering deployment in practice.

In order to bridge this gap, this paper proposes a very short-term wind power forecast method based on a VMD-XGBoost ensemble. Initially, the original power series is decomposed into IMFs. Next, for each IMF, an independent XGBoost model has been established with lagged features to capture temporal correlations better. To obtain the final power forecast, all IMFs' predictions are summed. The new approach is capable of providing long-term accuracy and is computationally efficient and interpretable, for being applicable in real-time dispatch in wind farms. With 3.3 years of measured data with 15-min resolution from a wind farm, we compare the proposed method thoroughly with the persistence model and raw XGBoost through five different forecast horizons (15 min, 30 min, 1 h, 2 h, and 4 h) to evaluate prediction performance and reveal the mechanisms behind predictions.

2 Data and Methods

2.1 Data Source and Preprocessing

The research uses data from the Supervisory Control and Data Acquisition (SCADA) system of a wind farm located in Shanxi province, China. The data are from October 1, 2019, to January 31, 2023 (around 3.3 years), and the sampling interval was 15 min. The name and location of the wind farm are commercially confidential. Data has been de-identified to protect the identity of the users.

The initial dataset consists of 110,057 records, which contain date, time, layer height (m), wind vane direction (°), wind speed (m/s), wind temperature (℃), wind vane pressure (kPa), relative humidity (RH), air density (Kg/m3), wind farm number, feeder number, short-term forecast power (MW), short-term reported power (MW), short-term manually corrected power (MW), ultra-short barn raw (MW), ultra-short term reported (MW), and actual power (MW).

The procedure of cleaning was: (1) We examined the validity of every field. Throughout the records, fields including layer height, temperature, pressure, humidity, and air density had a value of zero (due to unconnected or faulty sensors) and were removed. (2) The column "short-term manually corrected power" had a missing rate >80% and was removed. (3) We removed the other rows with fewer missing values. A total of 104,231 valid records were achieved. The final modeling features selected are: wind speed ("wind vane speed"), wind direction ("wind vane direction"), actual power and timestamp. From hereafter, we shall use "wind speed" and "wind direction" uniformly.

In order to remove abnormal points in the wind speed-power curve (e.g., shutdowns,

curtailments, or failures), we used the Pauta criterion (3σ) based on the theoretical power curve. First, we fitted a logistic function to the normal region with wind speed between 3 and 20 m/s and power greater than 0.1 MW:

$$P_{\text{fit}}(v) = \frac{L}{1+e^{-k(v-v_0)}} \tag{1}$$

where L=249.0 MW is the maximum power, and the fitting parameters are k=0.363, and v_0 =10.10 m/s. Based on the turbine specifications, the rated wind speed is v_{rated}=14.0 m/s, yielding a saturation power $P_{rated}=P_{\text{fit}}(v_{\text{rated}})$=200.4 MW. The cut-out wind speed is 25 m/s, above which turbines shut down. The theoretical power curve is defined as a piecewise function:

$$P_{\text{theory}}(v) = \begin{cases} P_{\text{fit}}(v), & v < v_{\text{rated}} \\ P_{\text{rated}}, & v_{\text{rated}} \le v < 25\ \text{m/s} \end{cases} \tag{2}$$

We calculated the residual r between each actual power and the theoretical value, and flagged points with $|\, r-\mu_r \,| > 3\sigma_r$ as anomalies. We also removed points with wind speed > 3 m/s and actual power <0.05 MW as shutdown anomalies. After cleaning, 102,382 valid samples remained, with a removal rate of 1.77%. All subsequent modeling was based on the cleaned complete dataset.

2.2 Variational Mode Decomposition (VMD)

2.2.1 Basic Principle of VMD

VMD decomposes the original signal $f(t)$ into K finite-bandwidth intrinsic mode functions (*IMFs*)$u_k(t)$, each centered around its own center frequencyω_k and assumed to be sparse in frequency. The decomposition is formulated as the following constrained variational problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad s.t. \sum_{k=1}^{K} u_k(t) = f(t) \tag{3}$$

By introducing the quadratic penalty factor α and the Lagrange multiplier λ(t), the problem is transformed into an unconstrained form and solved iteratively using the alternating direction method of multipliers (ADMM).

2.2.2 Selection of the Number of Modes K

K is the key parameter. Too small a K leads to under-decomposition (mode mixing), while too large a K yields spurious components. Using the center-frequency method to determine K: We ran VMD for K = 2, 3, . . . , 7 and then calculated the center frequency of the IMF. We observed how the highest-frequency IMF's center frequency changes with K; the K at which it stabilizes was taken as optimal. If two center frequencies become very close as K increases, that indicates over-decomposition. We set α=2000, τ=0, which are common choices in the literature[14]. The convergence tolerance was 10^{-7}. It is important to note that the center frequencies and bandwidths are determined solely from the training set. For the validation and test sets, these parameters are fixed; each IMF is extracted via Wiener filtering (i.e., band-pass filtering using the frequency support ranges from the training set) in a sliding-window manner with a window length equal to the entire historical training set length (i.e., using all past data up to the current time), strictly avoiding future information leakage.

2.3 Feature Engineering

2.3.1 Common Features

We extracted the following features from historical data (all lagged, no future leakage):

(1) Temporal features: hour (0–23), day of week (0–6), month (1–12), and weekend indicator (0/1), extracted from the timestamp to capture diurnal, weekly, and seasonal patterns.

(2) Wind direction encoding: angles (0–360 °) were transformed into sine and cosine components to remove circular discontinuity: $W_{dir\_sin}=\sin(\theta\ \pi/180)$， $W_{dir\_cos}=\cos(\theta\ \pi/180)$.

(3) Lag features: first- to sixth-order lags of wind speed and power (i.e., 15, 30, ..., 90 min earlier, corresponding to 1- to 6-step lags at 150-min resolution): P_{t-1} to P_{t-6} and V_{t-1} to V_{t-6}. Preliminary tests showed that increasing from 6 to 8 lags reduced 4-h RMSE by <0.5%, so we chose 6 as a trade-off.

(4) Rolling statistics: By calculating the wind speed's mean and standard deviation over the past hour (4 samples), we capture recent average and variability. We opted for a 1-hour window, as it is a typical inertial time scale and highly relevant for 2-4 h forecasts; preliminary tests with 2-hour windows yielded negligible gains (RMSE reduction 0.3% or less), so we kept 1 hour for simplicity.

In total, there are 20 common features: 4 temporal, 2 wind-direction, 6 power-lag, 6 wind-speed-lag, and 2 rolling statistics.

The target is the actual power at the future n_{step}-th sampling instant, with $n_{step} \in \{1,2,4,8,16\}$ corresponding to horizons of 15 min, 30 min, 1 h, 2 h, and 4 h, respectively. We train a separate XGBoost model for each horizon.

2.3.2 IMF-Specific Features (Only for VMD-XGBoost Model)

For each $IMFu_k$, we add its own lagged features to better capture its dynamics. For each k, the following features are constructed:

$$F_{IMF_k}=\{u_{k,\,t-1},\, u_{k,\,t-2},\, u_{k,\,t-3}\} \tag{4}$$

We chose a third-order lag because the autocorrelation of high-frequency IMFs decays quickly, and 45 min (3 steps) suffices to capture their main dynamics. Preliminary tests showed that increasing from 3 to 6 lags reduced 4-h RMSE by only 0.3% but doubled the feature size, so we kept 3 as a trade-off. Thus, the input dimension for the k_{th} IMF model is$|F_{common}|+ 3 = 23$.

2.4 Prediction Models

2.4.1 Raw XGBoost Model (Baseline)

XGBoost is a gradient-boosting ensemble that iteratively adds regression trees to fit residuals of previous predictions, progressively reducing error. Given training data $\{(x_i,y_i)\}$, the objective function is:

$$L = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega\,(f_t) \tag{5}$$

where l is the loss function (mean squared error), and $\Omega(f_t)=\gamma T+\frac{1}{2}\lambda\|\omega\|^2$ is the tree-complexity penalty. We determined hyperparameters via grid search: $n_{estimators}$=150, max_{depth}=3, learning rate=0.05, sub_{sample}=0.8, and $colsample_{bytree}$=0.8. The raw XGBoost uses F_{common} to directly predict the n_{step}-ahead power.

2.4.2 VMD-XGBoost Ensemble Model

(1) VMD decomposition: We apply VMD to the training set to obtain center frequencies ω_k and bandwidths for each IMF. For validation and test sets, we fix these parameters and extract each IMF via Wiener filtering (band-pass filtering using the frequency support from training) in a sliding-window manner using all available past data up to the current prediction time, which avoids any future leakage. Then, for each IMF, we construct its own lags (t-1, t-2, t-3) as additional inputs.

(2) Independent modeling of each IMF: For each IMF*uk*, we train a separate XGBoost model with inputs $F\text{common} \cup F_{IMF_k}$ and target $u_{k,t+n_{step}}$. Each model uses the same chronological split (70% training, 15% validation, 15% test) and the same hyperparameters as raw XGBoost.

(3) Prediction summation: For each test point, we sum all IMF predictions to obtain the final power forecast:

$$\hat{P}(t + n_{step}) = \sum_{k=1}^{K} \hat{u}_k\,(t + n_{step}) \tag{6}$$

2.4.3 Persistence Model (Strong Baseline)

The persistence model simply uses the current power P_t as the forecast for the k_{th} future step:$\widehat{P}_{t+k} = P_t$. Its physical basis is the continuity of the atmospheric boundary layer, wind-field inertia, and turbine mechanical inertia. The first-order autocorrelation of power is as high as 0.995 on the training set, so persistence is near-optimal for 15-min to 1-h forecasts. We use it as a strong baseline, which is a core design of this study.

2.4.4 Deep Learning Baselines (LSTM and TCN)

For additional comparison, we implemented two deep learning models: Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN). Both models were trained on the same common features (20 features) without IMF decomposition. The LSTM had 64 hidden units, 2 layers, and a learning rate of 0.001. The TCN used kernel size 3, 64 channels, and a learning rate of 0.001. Hyperparameters were selected via grid search. These models serve as representative deep learning baselines to benchmark the proposed ensemble method.

2.5 Experimental Setup

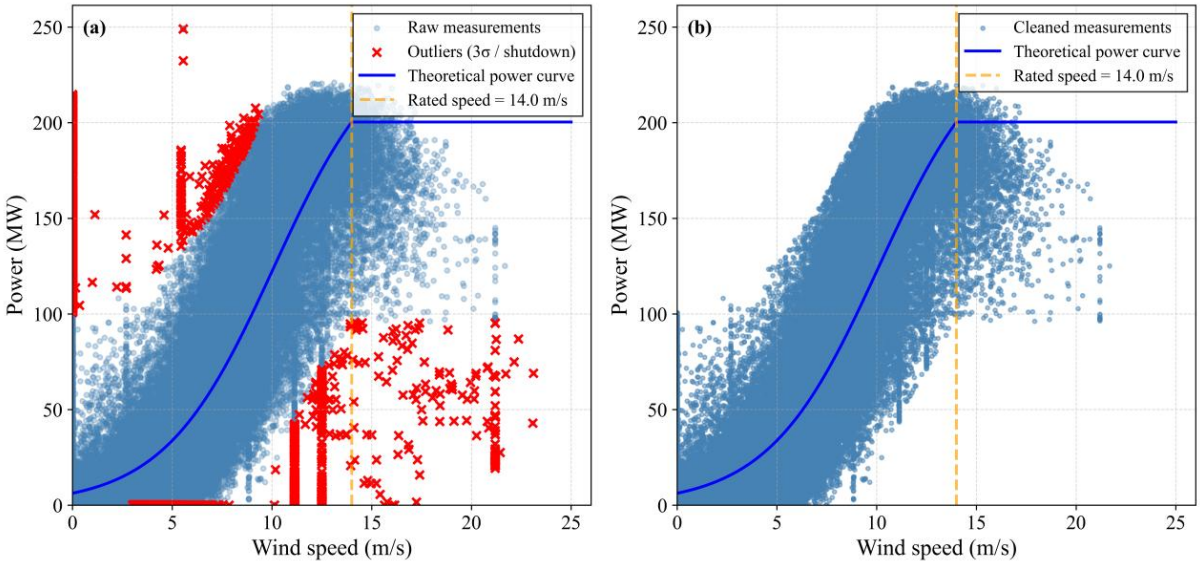
Five typical ultra-short-term horizons are considered: 15 min, 30 min, 1 h, 2 h, and 4 h, corresponding to *n*step=1, 2, 4, 8, and 16 (original 15-min resolution). All models are chronologically divided into training/validation/test sets at the same time (70%/15%/15%); as such, no future leakage is possible. The test set is utilized for the calculation of the evaluation metrics MAE, RMSE, and R^2.

3 Results

3.1 Wind Speed-Power Curve Cleaning Effect

The Pauta criterion was used to reject abnormal data. According to Figure 1(a), the wind speed-power scatter plot of our site shows many points below the power curve (power at 0) and scattered point at high wind speed region. Figure 1(b) illustrates the data that underwent cleaning. Subsequently, it follows the theoretical power curve more closely, with all points being much closer. The figure's red crosses (×) show the points deleted by the 3σ criterion

method. Of the previous 104,231 records, 1,849 outliers were dropped, leaving 102,382 valid records after cleaning the data. Outlier removal rate of 1.77%, greater than the theoretical value under normal distribution assumption (approximately 0.3%), is mainly caused by actual operational data's physical anomalies like power curtailment, shutdown, and equipment failure.



(a) Raw data (light blue points) and outliers (red ×); (b) Data after outlier removal and theoretical power curve.

Figure 1: Wind speed-power curve.

3.2 VMD Decomposition Results and Mode Number Selection

The original power series (15-min resolution, having 102,382 valid points) was decomposed with variational mode decomposition using mode number K=6 and penalty factor α=2000. Figure 2. Displays time-domain waveforms of the actual series and six IMF components (only the first 10,000 points are displayed for clarity purposes). It can be seen that the IMF1–IMF3 are those high-frequency oscillatory components that reflect rapid random power fluctuations. The IMF4–IMF6 on the other hand, are the low-frequency trend components, and these reflect the diurnal and seasonal variations. VMD is effective in decomposing a non-stationary power series into several approximately stationary power series, as indicated by the frequency separation of the components and the absence of the mode-mixing phenomenon, which lays a good foundation for subsequent XGBoost model learning.

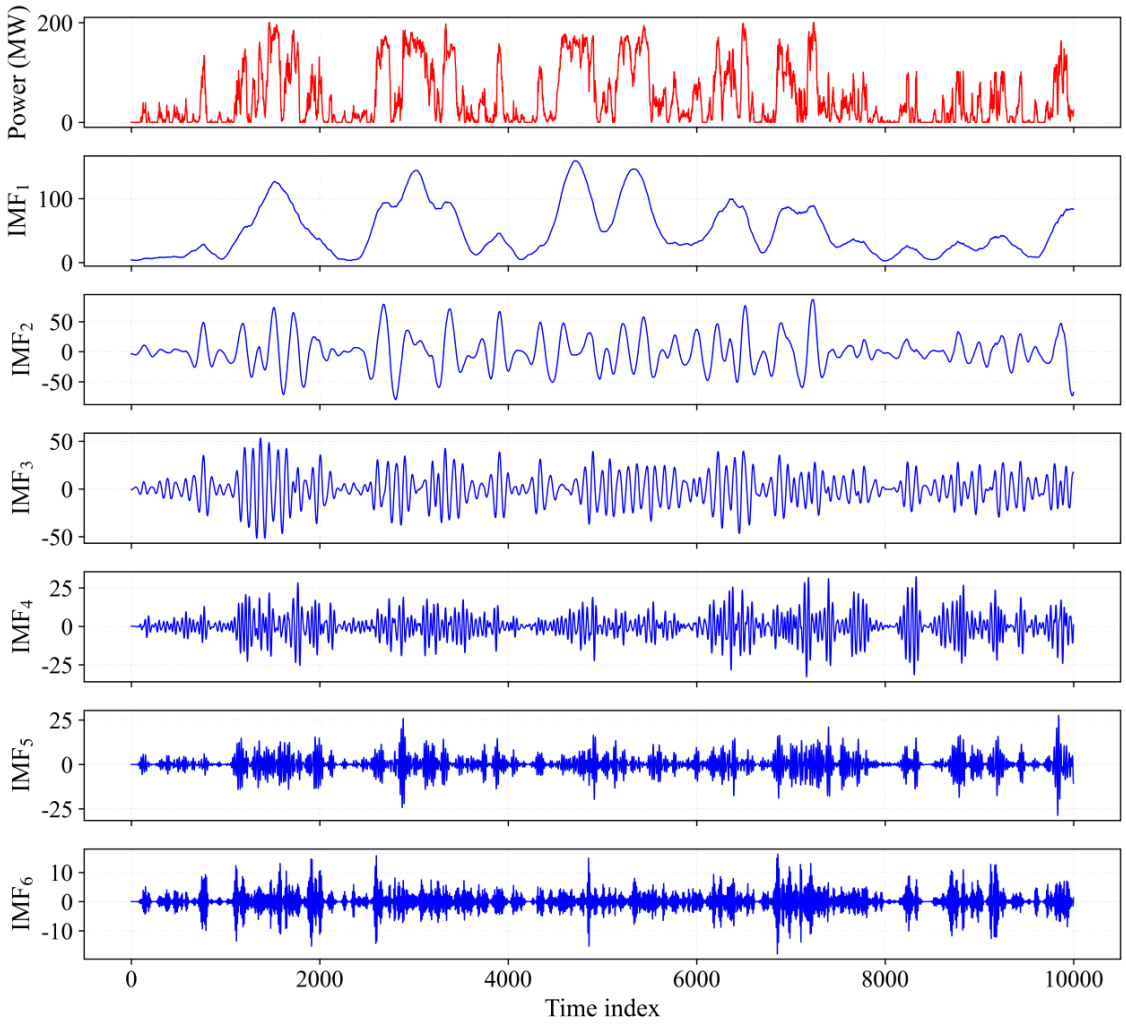


Figure 2: Original power series and six IMF components after VMD decomposition (only the first 10,000 points are shown).

To validate the impact of K on prediction performance, comparative experiments were conducted with K=4, 5, 6, and 7 (4-h forecast), with results shown in Table 1. K=6 achieved the best accuracy-efficiency balance. K=4 led to a significant decrease in R^2 due to under-decomposition, while K=7 exhibited over-decomposition with only marginal accuracy improvement but substantially increased computational cost. Therefore, K=6 was selected in this study.

Table 1: 4-h forecast performance comparison under different K values

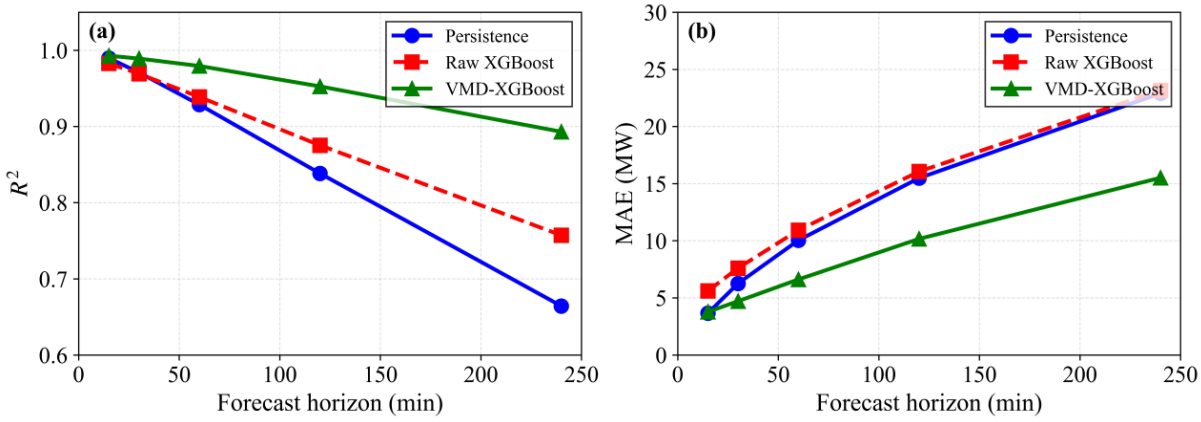
K	R^2	Training Time (s)	Remarks
4	0.8513	52	Mode mixing, high non-stationarity
5	0.8801	60	—
6	0.8922	68	Selected in this study
7	0.8945	85	R^2 improvement only 0.0023, time +25%

3.3 Model Performance Comparison Across Different Forecast Horizons

RMSE, MAE, and R^2 were computed for each model at the five horizons, with results shown in Table 2. Figure 3 presents the trends of R^2 and MAE with increasing forecast horizon. ① At

the 15-min horizon, the first-order autocorrelation coefficient of the power series is as high as 0.994, and the persistence model achieves extremely high performance (R^2=0.9897). Raw XGBoost's RMSE (8.669 MW) and R^2 (0.9830) are slightly inferior to the persistence model. VMD-XGBoost achieves the highest R^2 (0.9927) and the lowest RMSE (5.685 MW), but its MAE (3.840 MW) is higher than that of the persistence model (3.656 MW). This is because VMD decomposition and independent modeling introduce additional minor random fluctuations that are not completely canceled out upon summation. Although the persistence model occasionally exhibits large tracking lags, its errors at most time points are very small. The MAE difference is merely 0.184 MW, which is practically acceptable. ② At the 30-min and 1-h horizons, the persistence model's R^2 drops to 0.9709 and 0.9288, respectively, and VMD-XGBoost shows clear advantages: at 30 min, R^2=0.9883, with RMSE 29.2% lower than the persistence model; at 1 h, R^2 =0.9794, with MAE 39.2% lower than raw XGBoost. ③ At the 2-h and 4-h horizons, the persistence model's R^2 drops to 0.8384 and 0.6643, while VMD-XGBoost maintains high accuracy: at 2 h, R^2 = 0.9525, with RMSE 38.3% lower than raw XGBoost; at 4 h, R^2=0.8922, with RMSE and MAE reduced by 33.6% and 32.9%, respectively. These results demonstrate that VMD decomposition significantly enhances long-horizon prediction capability. ④ To provide a more comprehensive performance evaluation, complementary comparative experiments with LSTM and TCN on the same dataset were conducted (LSTM: 64 hidden units, 2 layers, learning rate 0.001; TCN: kernel size 3, 64 channels, learning rate 0.001). Both models used the same grid search strategy for hyperparameters, but their performance remained inferior. At the 4-h horizon, LSTM achieved R^2 = 0.7110 and RMSE = 35.731 MW; TCN achieved R^2=0.6984 and RMSE=36.522 MW, both considerably less accurate than VMD-XGBoost. Their training times were 12.4 times (approximately 843 s) and 8.2 times (approximately 560 s) that of the proposed method (68 s), respectively, verifying the comprehensive advantages of the proposed ensemble strategy in terms of both accuracy and efficiency.

Additionally, we analyzed the residual distribution of VMD-XGBoost at the 4-h horizon. The residuals have a mean of -0.21 MW and are approximately normally distributed (Shapiro-Wilk test *p*-value > 0.05), indicating unbiased predictions and supporting the error cancellation effect discussed later.

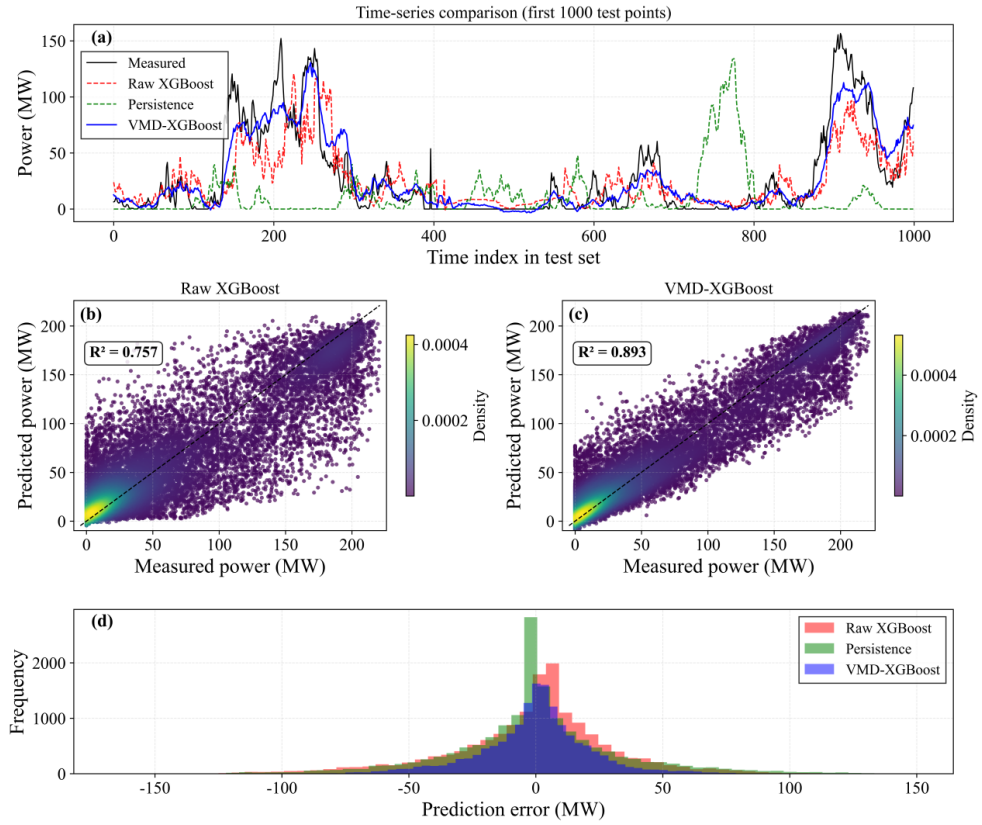


(a) R^2 vs. forecast horizon; (b) MAE vs. forecast horizon.

Figure 3: Comparison of R^2 and MAE across different forecast horizons for each model.

Table 2: Performance comparison of different models at each horizon

Horizon	Model	RMSE (MW)	MAE (MW)	R^2
15 min	Persistence	6.015	3.656	0.9897
	Raw XGBoost	8.669	5.599	0.9830
	VMD-XGBoost	5.685	3.840	0.9927
30 min	Persistence	10.138	6.262	0.9709
	Raw XGBoost	11.651	7.597	0.9693
	VMD-XGBoost	7.181	4.895	0.9883
1 h	Persistence	15.857	10.038	0.9288
	Raw XGBoost	16.381	10.891	0.9392
	VMD-XGBoost	9.539	6.622	0.9794
2 h	Persistence	23.882	15.486	0.8384
	Raw XGBoost	23.479	16.065	0.8752
	VMD-XGBoost	14.486	10.177	0.9525
4 h	Persistence	34.422	22.903	0.6643
	Raw XGBoost	32.847	23.199	0.7557
	VMD-XGBoost	21.815	15.572	0.8922
	LSTM	35.731	25.570	0.7110
	TCN	36.522	26.601	0.6984



(a) Time-series comparison (measured vs. predictions from each model); (b) Scatter plot of measured vs. predicted for raw XGBoost; (c) Scatter plot of measured vs. predicted for VMD-XGBoost; (d) Error distribution histograms of the three models.

Figure 4: Forecast performance comparison of different models at the 4-h horizon.

Figure 4 shows the actual prediction performance of each model at the 4-h horizon. From the time-series comparison in Figure 4(a), it can be observed that the persistence model (green dashed line) exhibits an obvious lag during periods of rapid power fluctuations. VMD-XGBoost (blue solid line) is closer to the measured power (black line). Scatter plots given in Figures 4(b) and 4(c) indicate that VMD-XGBoost achieves R2=0.892, which is much higher than the raw XGBoost which attained a score of 0.756. Figure 4(d) depicts the error distribution histogram, which confirms that the prediction errors generated by VMD-XGBoost are rather concentrated with fewer extreme errors.

3.4 Feature Importance Analysis

The gain-based importance of the model input feature was analyzed by using the XGBoost model for the IMF_1 high-frequency component. Figure 5 shows the normalized top 8 important features: IMF_1's own lag2 (0.2855), lag1 (0.2437), lag3 (0.1646), hour (0.0554), power lag2 (0.0458), power lag6 (0.0234), power lag1 (0.0223), and wind speed (0.0207). The three IMF lags' cumulative gain importance is around 69.38%, indicating that the high-frequency component's prediction highly depends on its own past values with extremely strong short-term autocorrelation. The term "hour" ranks fourth in importance among all temporal features with an importance of 0.055, which indicates that high-frequency fluctuations are also driven by diurnal cycles. The common features' contributions of power lags and wind speed are significantly lower.

Additionally, in order to comprehensively reveal the evolution pattern of the contribution of each frequency-band component to the prediction horizon, the contribution ratio of the prediction error of each IMF at three horizons, 15 min, 2 h, and 4 h (the squared prediction error of each IMF to the sum of the squared prediction errors of all IMFs). At the 15-min horizon, the error due to the high-frequency components (IMF_1–IMF_3) contributes to around 78% while low-frequency components (IMF_4–IMF_6) contribute to around 22%. After 2 hours, the high frequency share is 55% and the low frequency share is 45%. The high-frequency share further declines to 38% at the 4-h horizon, while the low-frequency share rises to 62%. This evolution quantitatively corroborates the transition mechanism from "inertia dependence" (driven by high-frequency own lags) to "local meteorological feature driving" (driven by low-frequency meteorological and power memory).

Owing to space limitations, this study focuses on presenting the feature importance of the high-frequency component. A similar analysis for the low-frequency component (IMF_6) (figure omitted) shows that the contributions of raw power lags and wind speed features in its prediction are significantly higher than those for the high-frequency component, consistent with the physical law that meteorological observation features play an enhanced driving role in long-horizon forecasting.

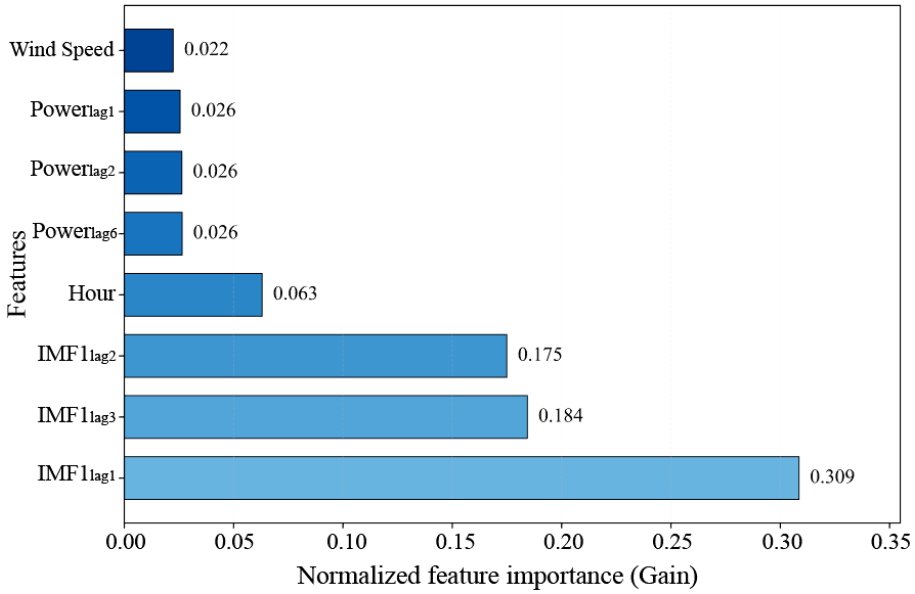


Figure 5: Feature importance of the XGBoost model for the high-frequency component IMF₁ (4-h forecast, based on gain).

3.5 Computational Efficiency

A single VMD decomposition execution takes approximately 2.1 s (Intel i7-10700 CPU). The serial training of the six IMF XGBoost models takes approximately 68 s in total (approximately 11–12 s per IMF). If parallel training is adopted (6-core parallel), the total time can be reduced to approximately 12 s. During the online prediction stage, after loading the six trained models, the inference time for a single 4-h forecast is less than 0.05 s, fully meeting the real-time requirements of ultra-short-term rolling forecasting.

4 Discussion

4.1 Sources of Performance Gain of the VMD-XGBoost Ensemble Model

The proposed model significantly outperforms the comparative models across all five horizons, with the most prominent improvements at the 2–4 h long horizons. The performance gains can be attributed to the following: ① Reduction of non-stationarity: VMD decomposes the original series into six relatively stationary IMFs so that each sub-model learns only the pattern of a specific frequency band, reducing learning difficulty. ②Independent modeling and feature enhancement: By modeling each IMF one by one and providing their own lagged features, we find that the prediction of high-frequency components heavily relies on their own lags (cumulative importance > 0.66), consistent with physical properties of high-frequency random perturbations. ③ Error cancellation effect of ensemble summation: The summation of each IMF's predictions is subject to random errors that cancel each other out. As demonstrated in Section 3.3, the residual mean was -0.21 MW and the residuals were approximately normally distributed, confirming unbiasedness and cancellation. ④ Horizon dependence of modal contributions: The examination presented in Section 3.4 demonstrates that the short-horizon gains predominantly arise from successful identification of high-frequency components. Furthermore, our long-horizon gains stem from the successful identification of low-frequency trend components.

4.2 Effect of VMD Mode Number K on Performance

The K=6 parameter is chosen from the sensitivity analysis as it gives good accuracy and moderate time. When K=4, due to the low K value, the non-stationarity of the score IMFs increases and effectively mixes the high-frequency and low-frequency components. Therefore, the XGBoost sub-models are relatively difficult to learn, leading to a decrease in the R^2 value by 0.041. When K=7, the R ²increases marginally by about 0.0023 while training time increases 25% over K=6 and the newly added mode's center frequency is very close to the highest-frequency mode's center frequency at K=6 (0.321 vs 0.314). Hence, it is a spurious component due to over-decomposition. Thus, K=6 provides an efficient and effective trade-off. The conclusion is aligned with the theoretical analysis of the center frequency method. Specifically, it is known that the best K corresponds with the point where the center frequency of the largest IMF level stabilizes.

4.3 Comparison with Existing Studies and Interpretability Value

The proposed approach has the following advantages over existing studies combining VMD and deep learning models like LSTM and GRU: ① High training efficiency: The decomposition of VMD only takes 2.1 s and serial training of the six XGBoost sub-models takes about 68 s, which can be reduced to 12 s with parallelization, much faster than the hours-long training of deep learning models. ② Strong interpretability: The gain importance analysis makes clear the physical mechanism where high-frequency components depend on their own lags and low-frequency components on meteorological information, which is a level of transparency that deep learning models do not offer. ③ Good adaptability to small and medium-sized datasets: With about 100,000 samples, the performance of XGBoost is excellent even in the case of small data. As seen in the LSTM/TCN results of Section 3.3, deep learning models are much less accurate than the proposed method on this data scale.

4.4 Limitations and Future Work

This study has the following limitations: ① Numerical weather prediction (NWP) data were not used. The inability of the model to capture future tendencies in meteorological aspects by using only wind speed and wind power on a historical basis confines the output to 6-h forecasting only. Incorporation of NWP data may enhance long-horizon accuracy. ② Single wind farm validation: Although the model performs well at the wind farm in Shanxi Province, the conclusions need to be validated at wind farms with more varied topographic and climatic conditions. ③ This study did not optimise VMD parameters (α, τ), selecting typical values of α=2000 and τ=0 for further tweaking. ④ Choosing a 1-hour rolling window for statistics on wind speed may not be optimal for 4-h forecasts. Although the experiments showed little additional gains from using longer windows, perhaps adaptive window selection can be explored. ⑤ The comparison with deep learning models remains insufficient: Although comparisons of LSTMs and TCNs have been added, comparisons with other advanced time-series models such as Transformer and N-HiTS are still needed.

Future research directions include: ① Incorporating NWP data: The proposed model uses numerical weather prediction information such as wind speed, wind direction, temperature and pressure as additional features to extend VMD-XGBoost to longer-horizon forecasting. ② Multi-wind-farm generalization validation: SCADA data from wind farms in different geographical surroundings to check model transferability. ③ Adaptive parameter optimization: An automatic search for the best K, α, and XGBoost hyperparameters using Bayesian

optimization or genetic algorithms. ④ More comprehensive deep learning comparison: Through systematic comparative analysis with state-of-the-art time-series frameworks such as Transformer and N-HiTS over the same dataset, further clarification of the accuracy boundaries of the proposed method.

4.5 Engineering Application Recommendations

The accompanying recommendations result from the present study's findings: ① The persistence model can be directly adopted for forecasting horizons ≤ 30 min, since its R2 remains above 0.97 and its cost is zero. ② For forecasting horizons ≥ 1 h, the VMD-XGBoost ensemble model is recommended, achieving an R^2 of 0.8922 at the 4-h horizon with inference time < 0.05 s, meeting real-time requirements. ③ A switching strategy can be adopted for forecasting systems covering the full range of horizons (15 min to 4 h): use the persistence model for ≤ 30 min and VMD-XGBoost for ≥ 1 h, achieving overall optimization of computational resources and accuracy. For newly commissioned wind farms, it is recommended to accumulate at least three months of historical data before deployment.

5 Conclusions

Based on 3.3 years of 15-min resolution measured data from a wind farm, this study proposes an ultra-short-term wind power forecasting method based on variational mode decomposition (VMD) and XGBoost ensemble, and rigorously compares it with the persistence model and raw XGBoost across five horizons: 15 min, 30 min, 1 h, 2 h, and 4 h. The main conclusions are as follows:

(1) VMD-XGBoost outperforms the comparative models at all horizons. At the 4-h horizon, R^2=0.8922, representing a relative improvement of 18.1% over raw XGBoost (0.7557), with RMSE and MAE reduced by 33.6% and 32.9%, respectively. At the 15-min horizon, R^2=0.9927 and RMSE=5.685 MW, which are better than the persistence model (0.9897, 6.015 MW). The comparison of LSTM and TCN further shows the efficiency of the proposed method.

(2) Predictions of high-frequency IMF's largely depend on their own lags with a cumulative importance of ~0.6938, while wind speed has little to no impact. In contrast, low-frequency IMFs depend more on temporal lags in energy and wind speed characteristics. The horizon-dependent evolution of modal error contributions quantitatively reveals a smooth transition from "inertia dependence" to "local meteorological feature driving": from 15 min to 4 h, high-frequency error share decreases from 78% to 38%, while low-frequency share increases from 22% to 62%.

(3) The approximate time for VMD decomposition is 2.1 s, parallel training of six XGBoost models is 12 s, and the inference of a single 4-h forecast is < 0.05 s, meeting real-time requirements. The gain-based importance analysis results in transparent and physically interpretable evidence.

(4) It is recommended to use a horizon-dependent model selection strategy, specifically, a persistence model for ≤ 30 min and VMD-XGBoost for ≥ 1 h. This solution does not need costly devices. Instead, it can be applied to regular servers. It will help reliably in the wind farm grid integration, dispatch and reserve capacity allocation.

In summary, the VMD-XGBoost ensemble method constructed by the researchers greatly enhances the accuracy of long-horizon ultra-short-term wind power forecasting and inherits its efficient training and strong interpretability from XGBoost, thus providing a practical technical path to the shift from "inertia dependence" to "local meteorological feature driving".

Author Contributions

Conceptualization, Hui Yuan and Minjie Chai; methodology, Hui Yuan and Siqing Xu; software, Hui Yuan and Jinsong Li; validation, Hui Yuan, Minjie Chai, and Jinwan Zheng; formal analysis, Hui Yuan; investigation, Minjie Chai; resources, Siqing Xu; data curation, Jinsong Li; writing—original draft preparation, Hui Yuan; writing—review and editing, Minjie Chai and Siqing Xu; visualization, Jinwan Zheng; supervision, Siqing Xu; project administration, Hui Yuan; funding acquisition, Hui Yuan. All authors have read and agreed to the published version of the manuscript.

Funding

The Science and Technology Project of State Grid Shanxi Electric Power Co., Ltd. supported this work (Grant No. 52053024003B).

Data Availability Statement

The dataset that is used in this study is commercially private and not publicly available. Upon reasonable request, the model code is available from the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Kumar K, Prabhakar P, Verma A, et al. Advancements in wind power forecasting: A comprehensive review of artificial intelligence-based approaches[J]. Multimedia Tools and Applications, 2025, 84(10): 8331-8360.

[2] Zhang D, Qian L, Mao B, et al. A data-driven design for fault detection of wind turbines using random forests and XGBoost[J]. IEEE Access, 2018, 6: 21020-21031.

[3] Wang Y S, Yang F, Qi Y S, et al. A Transformer-LSTM Network Enhanced by EEMD for Ultra-Short-Term Wind Power Forecasting[J]. Energy and AI, 2026, 23: 100682.

[4] Guo X, Zeng P, Xiong X, et al. Short-term wind power forecasting methods based on machine learning: A review and case study[J]. Energy Reports, 2025, 14: 3753-3782.

[5] Tuncar E A, Sağlam Ş, Oral B. A review of short-term wind power generation forecasting methods in recent technological trends[J]. Energy Reports, 2024, 12: 197-209.

[6] Song Z, Gu Y, Liu H, et al. Application of deep learning in wind, solar, and ocean energy: An analysis of prediction, optimization, and operation & maintenance[J]. Renewable and Sustainable Energy Reviews, 2026, 230: 116663.

[7] Yu G, Shen L, Dong Q, et al. Ultra-short-term wind power forecasting techniques: comparative analysis and future trends[J]. Frontiers in Energy Research, 2024, 11:

1345004.

[8] Xu T, Xie K G, Wang Y, et al. Ultra-short-term wind power forecasting based on TCN-Wpsformer hybrid model[J]. Power Automation Equipment, 2024, 8: 54-61.

[9] Xu W, Wang Y, Peng L, et al. A Regional Short-Term Wind Power Prediction Method Integrating DQN Error Correction with GCN-TCN-Transformer[J]. Processes, 2026, 14(8): 1275.

[10] Qian Z, Pei Y, Zareipour H, et al. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications[J]. Applied Energy, 2019, 235: 939-953.

[11] Zheng Q, Liu Y, Gu T. Wind power forecasting with a VMD-LSTM-Informer hybrid deep learning model[J]. Energy and AI, 2026, 23: 100765.

[12] Khan M A, Asif M, Alhussein M, et al. A dual-level hybrid machine learning model based on RFE-XGBoost and adaptive DBSCAN algorithm for multi-horizon wind energy forecasting[J]. Engineering Computations, 2026: 1-32.

[13] Zha W, Liu J, Li Y, et al. Ultra-short-term power forecast method for the wind farm based on feature selection and temporal convolution network[J]. ISA Transactions, 2022, 129: 405-414.

[14] Dragomiretskiy K, Zosso D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2013, 62(3): 531-544.

[15] Alam M M, Hossain M J, Habib M A, et al. Artificial intelligence integrated grid systems: Technologies, potential frameworks, challenges, and research directions[J]. Renewable and Sustainable Energy Reviews, 2025, 211: 115251.

[16] Zhou D, Jia Y, Liu G, et al. Research on an Ultra-Short-Term Wind Power Forecasting Model Based on Multi-Scale Decomposition and Fusion Framework[J]. Symmetry, 2026, 18(2): 253.

[17] Cai J, Wang Y. Short-term wind power prediction based on adaptive Stacking integrated learning model[J]. Physica Scripta, 2025, 100(6): 066004.

[18] Dong X, Zhang X, Yang M, et al. STDHL: Spatio-Temporal Dynamic Hypergraph Learning for Wind Power Forecasting[J]. arXiv, 2024, arXiv:2412.11393.