ORIGINAL RESEARCH

Wind Power Forecasting Based on Variational Mode Decomposition and Neural Hierarchical Interpolation

Xingru Ye[1,2] | **Ronghua Zhu**[1,3] | **Chenghong Gu**[2] | **Shufeng Dong**[4] | **Qiuyu Yan**[4] | **Zhisheng Tu**[1]

[1]Ocean College, Zhejiang University, Zhoushan, China

[2]Department of Electronic and Electrical Engineering, University of Bath, Bath, UK

[3]Yangjiang Offshore Wind Energy Laboratory, Yangjiang, China

[4]College of Electrical Engineering, Zhejiang University, Hangzhou, China

Correspondence
Ronghua Zhu
Email: zhu.richard@zju.edu.cn

Abstract

A method based on Variational Mode Decomposition (VMD) and Neural Hierarchical Interpolation for Time Series (N-HiTS/NHITS) is proposed to improve the forecast accuracy of wind power. Firstly, the VMD approach is adopted to decompose wind power signals into low-frequency and high-frequency components, which are then used as the input for the NHITS model. Compared with traditional forecast methods, this method based on VMD-NHITS can adapt to wind power's volatility and randomness and achieve better prediction results. Finally, the performance of the proposed forecast method is tested by using the wind power signal from an offshore wind farm in East China. Whether evaluated by point forecast metrics or probability forecast metrics, the proposed forecast method outperforms other forecasting methods in terms of prediction accuracy and reliability.

KEYWORDS

Wind Power, Ultra-short-term Forecast, Variational Mode Decomposition (VMD), Neural Hierarchical Interpolation

1 | INTRODUCTION

1.1 | Background and Motivation

Wind power's inherent volatility and randomness present significant challenges for grid stability. According to the technical and functional specifications outlined by the State Grid Corporation of China for wind farm grid connection, there is a stipulation that wind farms must integrate a wind power forecast system into their operations. This system is expected to exhibit ultra-short-term wind power forecast capabilities, ranging from 15 minutes to 4 hours. High accuracy and low computational time are essential benchmarks for assessing ultra-short-term wind power prediction models.

1.2 | Neural Networks in Wind Power Forecasting

Given the efficiency of neural networks in handling large-scale data, they have been widely used in predicting renewable energy systems. This includes classic single-method forecasts such as Convolution Neural Network (CNN) Acikgoz (2022), Temporal Convolution Network (TCN) Lin et al. (2021), Long Short Term Memory Neural Network (LSTM)Ko et al. (2021), Gated Recurrent Unit (GRU) Li et al. (2020), transformerWu et al. (2022), and their variants. Additionally, there has been a trend in recent years towards combining neural networks with other methods into multi-method forecast models. For instance, Flores et al. combined GRU with data augmentation for wind speed forecasting Flores et al. (2021). Another study by Li et al. presented a spatiotemporal directed graph CNN for multisite wind power prediction, highlighting the advantages of deep learning in wind energy forecasting Li et al. (2023). These advancements have made significant contributions to data identification, enhancing forecast resolution, and testing generalisation performance. However, they are still based on traditional neural

networks and are not easily interpretable and applicable to renewable energy practitioners. Thus, in recent years, time series introduced two novel interpretable neural methods, Neural Basis Expansion Analysis for Interpretable Time Series (N-BEATS/NBEATS) Olivares et al. (2023) and NHITS Challu et al. (2023) to improve the accuracy of forecasts. They emphasise learning different scales of the time series, resulting in improved performance. NBEATS differs from traditional approaches that decompose time series data into trend and seasonality components. Instead, it directly incorporates the decomposition of trend and seasonality into the model, enhancing the architecture to generate interpretable outputs without significantly sacrificing accuracy. NHITS is an extension of the NBEATS model, which can improve prediction accuracy and reduce computational cost. This wavelet-inspired algorithm combines the forecasts made at different time scales, considering both long-term and short-term impacts when generating the predictions, called hierarchical interpolation.

1.3 | Wind Power Decomposition

Time-series forecast methods such as NBEATS and NHITS gain insights into underlying patterns, trends, cycles, and seasonal effects. However, they do not provide much understanding of wind power's volatility and randomness characteristics. When it comes to ultra-short-term wind power forecasts, their capability to capture instantaneous changes is limited.

Therefore, SD approaches are introduced to solve the problem of ultra-short-term wind power forecasting. While low-pass filtering Sun et al. (2019), Liao et al. (2022), wavelet analysis Liu et al. (2019), and wavelet packet decomposition Dolatabadi et al. (2020) were previously common for wind power decomposition, techniques like EMD (Empirical Mode Decomposition) and its variants are now widely used. These methods excel in forecasting different frequency components of wind power due to their ability to handle nonlinear and nonstationary signals, as well as their effectiveness in multi-scale analysis. This makes them powerful tools for signal decomposition and feature extraction tasks. For instance, Shen et al. proposed a method using Ensemble Empirical Mode Decomposition (EEMD) and a Savitzky-Golay (SG) filter, with predictions made by LSTM and ARIMA models for wind speed Shen et al. (2021). Although these methods effectively capture the deep temporal features of wind speed time series, they still face challenges in mode mixing and noise robustness when dealing with real-world complex nonstationary signals. To further improve the accuracy of predictions and provide better control over the decomposition process, the introduction of VMD becomes particularly important. VMD reduces mode mixing and enhances noise robustness, offering significant advantages for handling complex nonstationary signals. For example, Han et al. improved multi-step wind power predictions by decomposing wind power data into three modes—long-term, fluctuation, and random using VMD and LSTM Han et al. (2019ab). The advantage of using VMD is that it allows for the pre-specification of mode number, which is especially valuable in applications where prior knowledge of the signal is available.

1.4 | Contributions

In this paper, we propose a novel VMD-NHITS-based forecast method drawing upon wind power frequency distribution and energy concentration characteristics for ultra-short-term wind power forecasting. It first utilises VMD to decompose wind power into low-frequency and high-frequency components, and then neural networks are employed for point prediction and probability prediction of each component. Finally, the predicted results for the components are combined to obtain the final wind power forecasts.

The main contributions of this paper are as follows:

1. A novel wind power forecast method based on VMD-NHITS is presented, which provides deliberate intervention in understanding wind power data by using the VMD approach before feeding data into a neural network for prediction. This intervention expedites the optimization procedure, thus enhancing the model's forecast ability.
2. Unlike relying only on a single forecast method, this method integrates low-frequency component point predictions with high-frequency component probability predictions to get final prediction results. This integration reduces the uncertainty in wind power forecasts, which can help facilitate more effective scheduling plans and standby strategies and enhance the overall stability and reliability of the power grid.

This paper is organised as follows. Section 2 describes the problem and the limitations of only using the NHITS method. In Section 3, the proposed forecast method based on VMD-NHITS is elaborated in detail. Section 4 compares the proposed novel method with previous methods from different evaluation aspects. The conclusion including work and prospects is presented in Section 5.

2 | PROBLEM STATEMENT

As the installed capacity of renewable energy sources continues to increase, the evaluation criteria for power forecasting are gradually becoming stricter. Taking the offshore wind farm in this paper as an example, the generation management and implementation rules in this area stipulate that when the ultra-short-term power forecasting accuracy can not reach 85% by daily assessment, the penalty fee is:

$$F = (0.85 - A) \times C \times 0.2 \times 0.25 \times P_r$$

where F represents the penalty fee, A is the accuracy which is calculated using the Root Mean Square Error (RMSE), C is the installed capacity, 0.25 is the technical management assessment coefficient for this area, and P_r represents the highest approved grid-connected electricity price of the generation. Improving the prediction accuracy of wind power can enhance the efficiency of wind farms and reduce the cost of standby capacity on the grid side, which has significant social and economic benefits. NHITS has been well used to increase the interpretability of the

neural network structure and improve the prediction accuracy and computational efficiency. However, this approach essentially obtains the components of different frequencies in the time series by changing the sampling rate (Explained in detail in Section 3). Thus, in order to improve the accuracy, this paper proposes to use the VMD approach for time series decomposition before inputting the training data into the neural network. The time series is decomposed into low-frequency and high-frequency sequences in advance, which allows subsequent neural layers to better extract features and improve accuracy.

3 | PROPOSED ARCHITECTURE AND METHODOLOGY

An overview of the proposed method consists of four steps. First, the module uses VMD to decompose wind farm output power into low-frequency and high-frequency components. Then, it is to train an NHITS model with low-frequency component inputs to obtain point predictions for the low-frequency components. Meanwhile, it is to train another NHITS model by using high-frequency component inputs to generate probability predictions for high-frequency components. Finally, the two components are aggregated to obtain the probability prediction of the final wind power. The procedure of the method is illustrated in Figure 1.
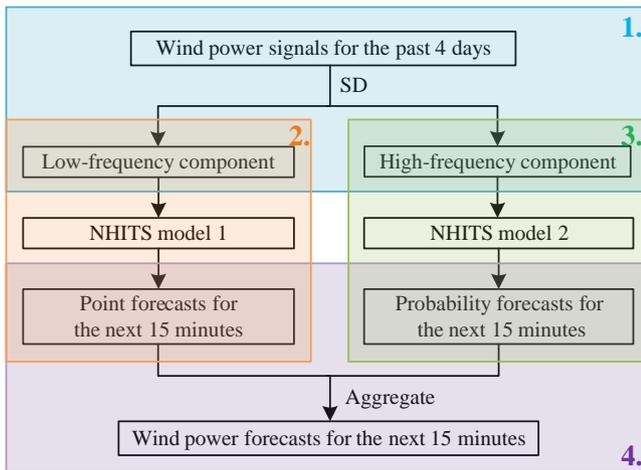
1. **Wind Power Decomposition Module**
2. **Low-frequency Component Point Forecast Module**
3. **High-frequency Component Probability Forecast Module**
4. **Forecasts Output Module**

FIGURE 1 Four steps of the forecast method.

3.1 | Wind Power Decomposition Module

Wind speed influences wind power generation primarily. The slow variation in wind speed causes the low-frequency component of wind speed,

while the instantaneous changes in wind speed and irregular wind flow induce high-frequency components. Therefore, the energy of wind power is dominated by low-frequency components. The amplitude-frequency characteristic curve of the offshore wind farm output is shown in Figure 2.

This figure confirms that wind energy is mainly dominated by low-frequency components, specifically with a remarkable concentration of 99.77% within the 0 to 0.2 mHz range. Therefore, wind power can be decomposed into low-frequency and high-frequency components, which can be processed separately. In this paper, VMD, with its advantages in accuracy, noise robustness and computational efficiency Eriksen and ur Rehman (2023), is employed as the SD approach. In essence, VMD is a multiple adaptive Wiener filter bank and the optimisation problem of VMD can be expressed as an energy minimisation problem, the goal of which is to find a set of IMFs so that their energies are dispersed as much as possible across different frequency ranges while maintaining the overall structure of the signal. Specifically, the optimisation problem of VMD can be expressed as follows:

$$\min_{\mathbf{u},\mathbf{h}} \sum_{k=1}^{K} \left\| x - \sum_{k=1}^{K} u_k * q_k \right\|_2^2 + \lambda \sum_{k=1}^{K-1} \|f_k - f_{k+1}\|_2^2$$

where x is the given time series; $\mathbf{u} = [u_1, u_2, \ldots, u_K]$ represents a set of Intrinsic Mode Functions (IMFs), K is the number of IMFs, u_k is the signal of the k-th IMF; $\mathbf{q} = [q_1, q_2, \ldots, q_K]$ is a set of orthogonal filters, where q_k is the filter corresponding to the k-th IMF; λ is a regularization parameter balancing the trade-off between data fitting and smoothing; f_k is the frequency of the k-th IMF obtained through the Hilbert transform.

When $K=2$, $\mathbf{P}_k$ is decomposed into two components, one is the low-frequency component $\mathbf{P}_1$ and the other is the high-frequency component $\mathbf{P}_2$.

3.2 | Low-frequency Component Point Forecast Module

In the forecasting model, a sliding window is used to form a multi-dimensional sample. The sliding window width is represented as W,
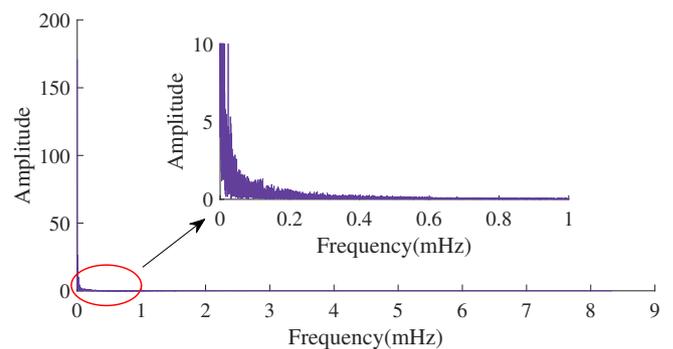

FIGURE 2 Amplitude-frequency characteristics of wind power.

encompassing a lookback period horizon of L and a forecast period horizon of H. Instead of applying VMD to the overall dataset, we perform VMD individually on the training data for each prediction model. This ensures that temporal information from the validation set is not leaked during the decomposition process, thereby maintaining the integrity of our time series prediction. In every forecasting model, wind power is a one-dimensional time series, so it can be denoted as $\mathbf{P}^{t-L|t+H}$. This window will move forward along the historical wind power series by a time step, forming a sample with each slide. The sliding window of signals is shown in Figure 3. The low-frequency component is relatively stable

FIGURE 3 Sliding window of signals.

and predictable. In such cases, point forecasts can effectively capture the expected future value without the need for additional complexity or uncertainty measures and Huber Loss is used as the loss function. Since point forecast does not involve probability distributions or ranges, the calculation process is relatively simple, resulting in higher computational efficiency. The structure of NHITS to handle the low-frequency component is depicted in Figure 4.
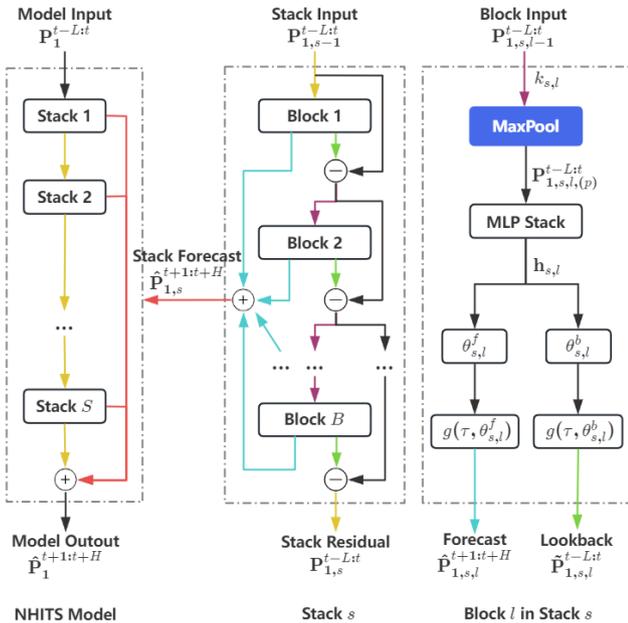

FIGURE 4 The structure of NHITS.

The input component $\mathbf{P}_1^{t-L:t}$ is divided into sub-time series with different periods by using different expressiveness ratios in different stacks. A low expressiveness ratio captures long-term trends, while a high expressiveness ratio captures short-term fluctuations. With the layered approach, NHITS can simultaneously consider more detailed low-frequency information and integrate this information into the forecasting process.

The stack is implemented by several blocks. The purpose of the stack being divided into blocks is to learn the basis functions of different characteristics. The lookback output is subtracted from the input of the next block so that the next block can analyse the remaining signal more intently. It amplifies the attention of the next block to signals outside the frequency band that have already been processed. All of the block forecast outputs are combined to form the stack forecast.

The Block l in Stack s uses a MaxPool layer with kernel size $k_{s,l}$ to cut small-time-scale components from the previous block residual $\mathbf{P}_{1,s,l-1}^{t-L:t}$, which can be indicated as

$$\mathbf{P}_{1,s,l,(p)}^{t-L:t} = \mathsf{MaxPool}\left(\mathbf{P}_{1,s,l-1}^{t-L:t}, k_{s,l}\right)$$

When $k_{s,l}$ is increased, it potentially results in a shorter input for the MLP $\mathbf{P}_{1,s,l,(p)}^{t-L:t}$. It is called multi-rate signal sampling.

By using non-linear regression with a hidden vector, symbolized as $\mathbf{h}_{s,l}$, the model gets forward and lookback interpolation MLP coefficients $\theta_{s,l}^f$ and $\theta_{s,l}^b$,

$$\mathbf{h}_{s,l} = \mathsf{MLP}_{s,l}\left(\mathbf{P}_{1,s,l,(p)}^{t-L:t}\right)$$

$$\theta_{s,l}^f = \mathsf{LINEAR}^f\left(\mathbf{h}_{s,l}\right)$$

$$\theta_{s,l}^b = \mathsf{LINEAR}^b\left(\mathbf{h}_{s,l}\right)$$

Since downsampling has been realized through the MaxPool layer, the forecast output will be less than the target Horizon, so a hierarchical interpolation is required on the prediction results to make the forecast sequence length reach H. Here the method uses the linear interpolator g, which is defined as

$$g(\tau, \theta) = \theta\left[t_1\right] + \left(\frac{\theta\left[t_2\right] - \theta\left[t_1\right]}{t_2 - t_1}\right)(\tau - t_1), t_1 = \arg \min_{t \in \mathcal{T}: t \leq \tau} \tau - t, t_2 = t_1 + 1/r_l$$

The Block l produces forecast output $\hat{\mathbf{P}}_{1,s,l}^{t+1,t+H}$ and lookback output $\widetilde{\mathbf{P}}_{1,s,l}^{t-L:t}$. The input of the next block $l+1$, described as $\mathbf{P}_{1,s,l}^{t-L:t}$, derives from $\mathbf{P}_{1,s,l-1}^{t-L:t}$ and $\widetilde{\mathbf{P}}_{1,s,l}^{t-L:t}$:

$$\mathbf{P}_{1,s,l}^{t-L:t} = \mathbf{P}_{1,s,l-1}^{t-L:t} - \widetilde{\mathbf{P}}_{1,s,l}^{t-L:t}$$

The forecast output of Stack s is assembled by summing the B block forecasts,

$$\hat{\mathbf{P}}_{1,s}^{t+1,t+H} = \sum_{l=1}^{B} \hat{\mathbf{P}}_{1,s,l}^{t+1,t+H}$$

Finally, the model output is assembled by summing the S stack forecasts,

$$\hat{\mathbf{P}}_1^{t+1,t+H} = \sum_{s=1}^{S} \hat{\mathbf{P}}_{1,s}^{t+1,t+H}$$

3.3 | High-frequency Component Probability Forecast Module

It has been mentioned that the instantaneous changes in wind speed and irregular wind flow induce the high-frequency component of wind power. This results in wind power having higher randomness and unpredictability, leading to significant errors in point forecasts. Therefore, a probability forecast is adopted to address this issue. The high-frequency component probability forecast module is almost the same as the low-frequency component point forecast module except for utilising Multi-Quantile Loss as the loss function.

3.4 | Forecasts Output Module

After finishing training the point prediction model and the probability model, the forecasts of low-frequency and high-frequency components an be represented as $\hat{P}_1, \hat{P}_{2,lo}, \hat{P}_{2,up}, \hat{P}_{2,m}$, where $\hat{P}_1$ is the point forecast for the low-frequency component, $\hat{P}_{2,lo}$ and $\hat{P}_{2,up}$ is the lower and upper bounds of the high-frequency component forecasts, and $\hat{P}_{2,m}$ is the median of the high-frequency component forecasts. The last step is to aggregate the point forecast of low-frequency components with the probability forecast of high-frequency components to obtain the final prediction results, including lower bounds $\hat{P}_{lo}$, upper bounds $\hat{P}_{up}$ and the median $\hat{P}_m$. Finally, this paper introduces five metrics for a comprehensive evaluation, two of which are metrics for point forecast, and the others for probability forecast Wan et al. (2014):

1. Calculate the Mean Squared Error (MSE) and Mean Absolute Error (MAE).
2. Calculate the Average Coverage Error (ACE), Prediction Interval Averaged Width (PIAW) and Winkler Score (WS).

In the evaluation, the smaller the metrics, the better the performance. Meanwhile, the Wilcoxon Signed-rank Test is also adopted to evaluate the statistical significance of the differences in model performance, with a significance level typically set at 0.05. The framework of the method is shown in Figure 5.

4 | CASE STUDY AND ANALYSIS

4.1 | Dataset Preparation and Parameters Settings

The wind power and SCADA data used in this paper from January to May 2022 were collected from an offshore wind farm in East China. The wind farm consists of 72 units of 4MW wind turbines, 1 unit of 6MW wind turbine, and 1 unit of 7MW wind turbine, with a total installed capacity of approximately 300MW. The timestamps of the dataset start from "2022-01-01 11:10:00" and end with "2022-05-31 10:45:00", containing 207022 sampled points at a temporal resolution of 1 min. All the
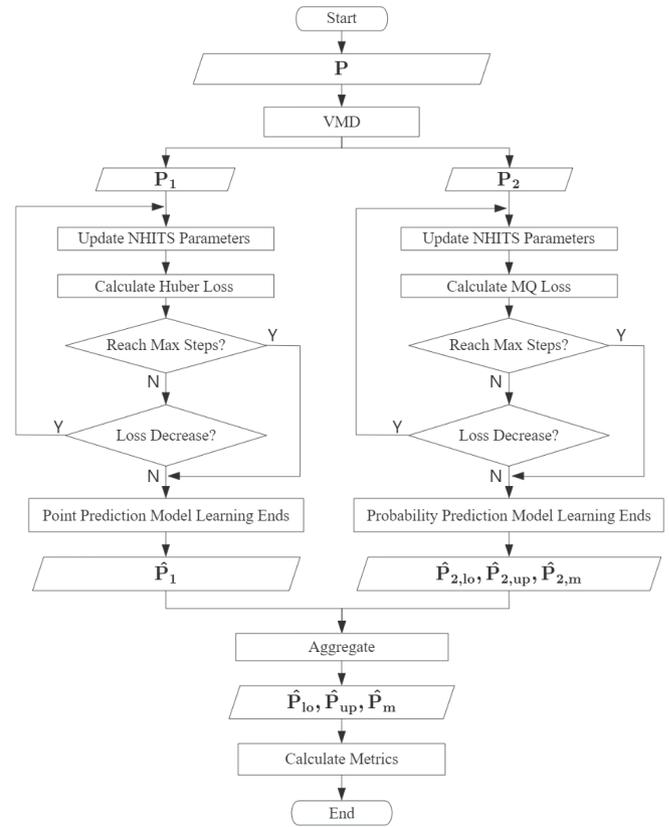

F I G U R E 5 Framework of the method.

SCADA data of the wind turbines in the wind farm are input into the neural network for training and testing. In that case, it will require a large amount of parameter calculation and consume a significant amount of computing resources. Therefore, based on the spatial layout of the wind turbines in the offshore wind farm, this paper selectively retains a part of turbines' wind speed data. Due to missing data for turbines 1-32 and a malfunction in turbine 54, the data from turbines 39, 45, 53, 55, 56, 60, 64, 68, and 72, a total of nine turbines, are chosen for processing. The layout of the wind turbines in the offshore wind farm is shown in Figure 6.

This paper implements the experiments on a 64-bit PC platform with an Intel Core-i7 CPU at 3.60GHz and an NVIDIA GeForce GTX 1650 GPU to deal with all models. The overall model training, validation, and testing are implemented on the PyTorch deep learning framework by Python 3.9.

In this work, a neural network model is periodically trained every 15 minutes. The model is fed with scaled data from the last 4 days for training. Here, the forecast period horizon H is set to 15 minutes and the lookback period L is set to 2 days. The length of testing data is the same as the forecast period H. The model's maximum step is set to 1000, and an early stopping mechanism is implemented with a patience setting of 2 steps during the model's training. The point forecast of the low-frequency component and the probability forecast of the high-frequency component will be superimposed after respective inverse
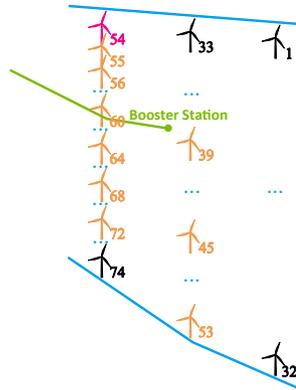
T A B L E 1   Evaluation of testing results with or without SCADA data under different prediction methods.

Data Input	Model	MSE	MAE	ACE	PIAW	WS	Wilcoxon Signed-rank Test p-value	$\alpha = 0.05$ Pass Y/N
Without SCADA	TCN	3834.23	43.11	120.79	118.55	123.04	2.43e-1	Y
	GRU	216.44	10.79	48.92	47.76	50.08	3.29e-7	N
	NBEATS	216.55	11.14	80.17	79.26	81.03	4.39e-1	Y
	NHITS	**166.41**	**9.54**	39.40	38.50	40.30	3.58e-1	Y
With SCADA	TCN	1087.18	26.04	62.22	52.71	71.72	3.88e-66	N
	GRU	1016.69	24.35	49.23	41.91	56.54	3.48e-9	N
	NBEATS	219.82	11.27	79.96	78.89	81.02	7.75e-1	Y
	NHITS	170.43	9.72	**37.05**	**36.19**	**37.92**	1.63e-1	Y

T A B L E 2   Metric improvement of each model compared by TCN without SCADA.

Data Input	Model	MSE	MAE	ACE	PIAW	WS
Without SCADA	TCN	0.00%	0.00%	0.00%	0.00%	0.00%
	GRU	94.36%	74.97%	59.50%	59.71%	59.30%
	NBEATS	94.35%	74.16%	33.63%	33.14%	34.14%
	NHITS	**95.66%**	**77.87%**	67.38%	67.52%	67.25%
With SCADA	TCN	71.65%	39.60%	48.49%	55.54%	41.71%
	GRU	73.48%	43.52%	59.24%	64.65%	54.05%
	NBEATS	94.27%	73.86%	33.80%	33.45%	34.15%
	NHITS	95.56%	77.45%	**69.33%**	**69.47%**	**69.18%**

without using SCADA data is still quite close to NHITS, but introducing SCADA data during model training worsens its performance.

Since the introduction of SCADA data in this study cannot stably and effectively improve the forecast accuracy and causes a large amount of computing resource consumption, the follow-up study excludes SCADA data, relying solely on wind power data. Moreover, since NHITS is an evolution of NBEATS and has been proven to outperform NBEATS, the follow-up study excludes NBEATS to reduce redundant work.

4.3.2 | Compare SD approach combined neural forecast models

In addition to the previously mentioned SD approaches, a 5th-order Butterworth low-pass filter is also used as a comparison model, separating low-frequency and high-frequency components with a cutoff frequency of 0.2 mHz. The prediction results of forecasting methods based on Filter-TCN, Filter-GRU, Filter-NHITS, EMD-TCN, EMD-GRU, EMD-NHITS, VMD-TCN, VMD-GRU, and VMD-NHITS are shown in Figure 9. As observed from Figure 9, VMD-NHITS exhibits the closest alignment with the original wind power data, especially in the peak and valley. And it has the narrowest bandwidth in probability prediction.

The evaluation metrics are presented in Table 3. By comparing Table 1 and Table 3, we can conclude that after VMD, the final aggregated results outperform the non-decomposed approach in terms of probabilistic prediction metrics, regardless of which neural network is used for point prediction on the low-frequency components and probabilistic prediction on the high-frequency components. What's more,

Table 3 also indicates that VMD-NHITS does not show significant improvement in point forecasting metrics compared to Filter-NHITS but shows the minimum values for the probability prediction evaluation metrics, suggesting its optimal probability prediction performance. The rough use of EMD in this experiment does not help improve the accuracy of the prediction. As for the significance test, only Filter-TCN, Filter-NHITS and VMD-NHITS pass the Wilcoxon Signed-rank Test (α=0.05). By contrast, all the failures of EMD in significance testing confirm that VMD of wind power is more adaptive to its characteristics than EMD.

The metric improvement compared to TCN without SCADA is shown in Table 4 and a radar chart displaying the improvement is shown in Figure 10. The VMD-NHITS model has the largest coverage area and the most uniform shape, indicating its superior overall performance. Overall, these charts showcase that the VMD-NHITS-based method for wind power forecasting is the most accurate and precise, exhibiting minimal prediction uncertainty, which makes it highly suitable for wind power prediction in wind farms.

5 | CONCLUSIONS

This work presents a novel VMD-NHITS-based wind power forecast method. The method uses the VMD approach to decompose wind power into low-frequency and high-frequency components, then uses NHITS models to produce point and probability forecasts separately and add up the forecasts to get the final prediction results. After being tested on the dataset and evaluated by all the metrics, the VMD-NHITS-based

TABLE 3 Testing results using different SD approach combined methods.

SD approach	Model	MSE	MAE	ACE	PIAW	WS	Wilcoxon Signed-rank Test p-value	$\alpha = 0.05$ Pass Y/N
	TCN	2747.66	36.03	59.47	39.45	79.49	1.62e-1	Y
Filter	GRU	575.20	16.95	58.82	56.37	61.26	4.77e-28	N
	NHITS	**171.51**	**9.91**	50.03	49.32	50.73	7.17e-1	Y
	TCN	2049.82	34.39	56.00	35.38	76.61	2.79e-2	N
EMD	GRU	695.59	20.46	40.99	30.32	51.66	2.30e-6	N
	NHITS	580.91	18.10	41.20	33.67	48.73	9.37e-6	N
	TCN	1754.67	27.48	44.43	25.36	63.49	8.15e-9	N
VMD	GRU	568.71	17.07	44.79	38.74	50.84	1.24e-22	N
	NHITS	227.97	11.31	**27.18**	**23.19**	**31.18**	5.82e-1	Y

TABLE 4 Metric improvement using different SD approach combined methods compared by TCN without SCADA.

SD approach	Model	MSE	MAE	ACE	PIAW	WS
	TCN	28.34%	16.42%	50.77%	66.72%	35.39%
Filter	GRU	85.00%	60.68%	51.30%	52.45%	50.21%
	NHITS	**95.53%**	**77.01%**	58.58%	58.40%	58.77%
	TCN	46.54%	20.23%	53.64%	70.16%	37.74%
EMD	GRU	81.86%	52.54%	66.07%	74.42%	58.01%
	NHITS	84.85%	58.01%	65.89%	71.60%	60.39%
	TCN	54.24%	36.26%	63.22%	78.61%	48.40%
VMD	GRU	85.17%	60.40%	62.92%	67.32%	58.68%
	NHITS	94.05%	73.76%	**77.50%**	**80.44%**	**74.66%**

forecast method has been proven to possess the ability to accurately forecast the wind power, which will be beneficial and meaningful for providing a decision-making basis for generator set control, automatic power generation control, standby, and other auxiliary service management, real-time economic scheduling, real-time safety analysis, congestion management, power spot market, renewable energy and energy storage collaborative control, etc.

The limitations of the research are lacking more detailed wind farm geographical information and feature extraction of SCADA data, and the power prediction is performed only for typical periods without considering extreme weather conditions. Thus, further research can focus on the following areas:

1. Integration of geographical Information: Incorporate detailed geographical information of wind farms as external variables into the model to further enhance forecasting accuracy.
2. Effective feature extraction from SCADA Data: Explore more effective feature extraction methods to improve the utilization efficiency of SCADA data in wind power forecasting.
3. Enhancing model adaptability: Study the model's adaptability under different climatic conditions and geographical locations to improve its generalisation capability.

AUTHOR CONTRIBUTIONS

Xingru Ye: Conceptualisation, methodology, software, writing – original draft. Ronghua Zhu: Funding acquisition, project administration, resources, supervision. Chenghong Gu: Validation, supervision, writing – review & editing. Shufeng Dong: Visualization, supervision, writing – review & editing. Qiuyu Yan: Investigation, data curation. Zhisheng Tu: Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

REFERENCES

Acikgoz, H. (2022) A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting. *Applied Energy*, 305. doi:10.1016/j.apenergy.2021.117912.

Bai, S., Kolter, J.Z. & Koltun, V. (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,.
URL http://arxiv.org/abs/1803.01271

Challu, C., Olivares, K.G., Oreshkin, B.N., Ramirez, F.G., Mergenthaler-Canseco, M. & Dubrawski, A. Nhits: Neural hierarchical interpolation for time series forecasting. Vol. 37, *2023*.

Dolatabadi, A., Abdeltawab, H. & Mohamed, Y.A.R.I. (2020) Hybrid deep learning-based model for wind speed forecasting based on dwpt and bidirectional lstm network. *IEEE Access*,.doi:10.1109/ACCESS.2020.3047077.

Eriksen, T. & ur Rehman, N. (2023) Data-driven nonstationary signal decomposition approaches: a comparative analysis. *Scientific*
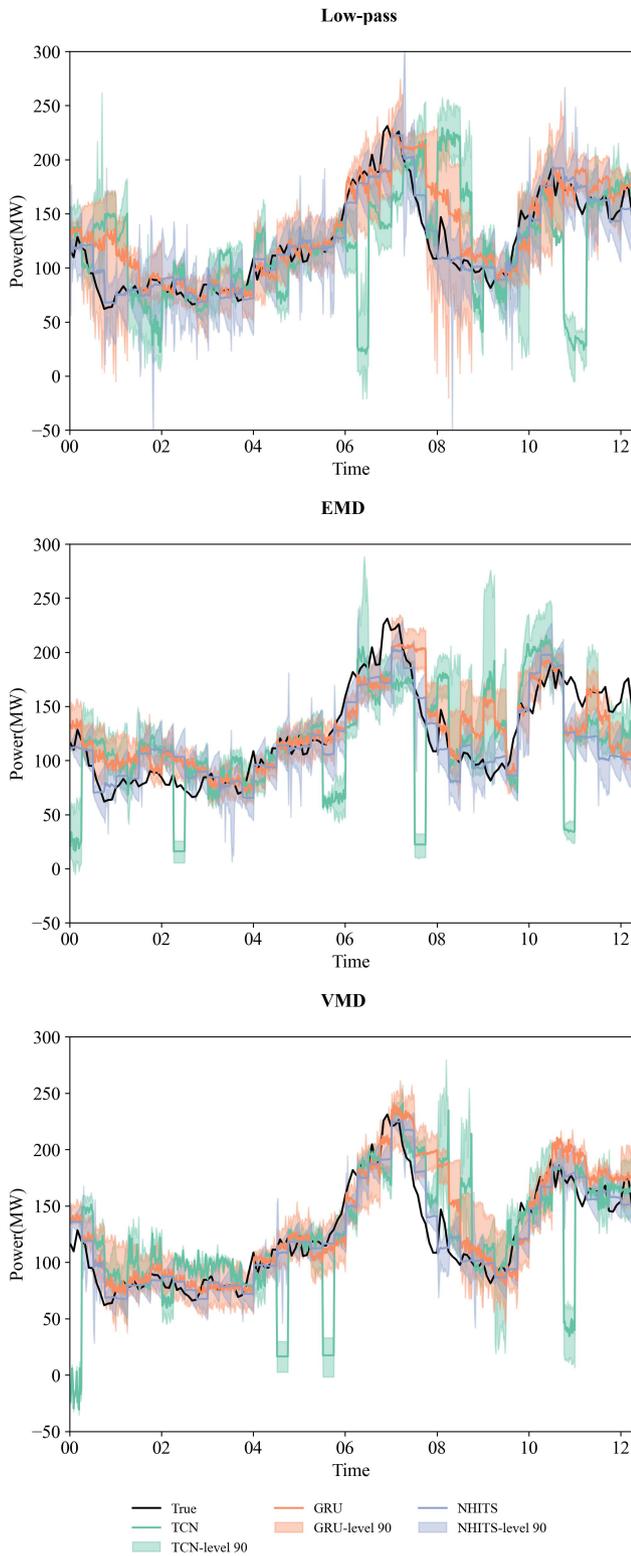
Low-pass

EMD

VMD

FIGURE 9 Comparison of results with different SD approach combined methods.
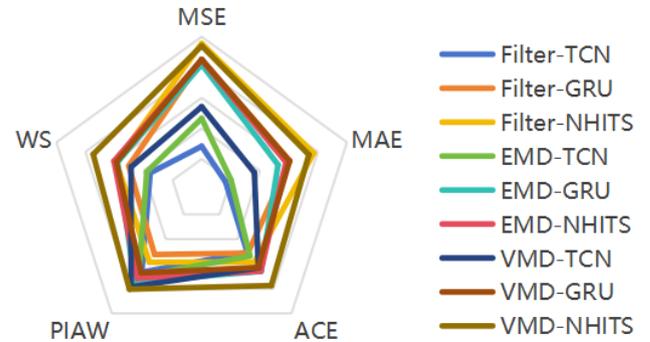

FIGURE 10 Radar chart of the metric improvement for different SD approach combined methods.

Flores, A., Tito-Chura, H. & Yana-Mamani, V. (2021) An ensemble gru approach for wind speed forecasting with data augmentation. *International Journal of Advanced Computer Science and Applications*, 12. doi:10.14569/IJACSA.2021.0120666.

Han, L., Jing, H., Zhang, R. & Gao, Z. (2019) Wind power forecast based on improved long short term memory network. *Energy*, 189. doi:10.1016/j.energy.2019.116300.

Han, L., Zhang, R., Wang, X., Bao, A. & Jing, H. (2019) Multi-step wind power forecast based on vmd-lstm. *IET Renewable Power Generation*, 13. doi:10.1049/iet-rpg.2018.5781.

Ko, M.S., Lee, K., Kim, J.K., Hong, C.W., Dong, Z.Y. & Hur, K. (2021) Deep concatenated residual network with bidirectional lstm for one-hour-ahead wind power forecasting. *IEEE Transactions on Sustainable Energy*, 12. doi:10.1109/TSTE.2020.3043884.

Li, C., Tang, G., Xue, X., Saeed, A. & Hu, X. (2020) Short-term wind speed interval prediction based on ensemble gru model. *IEEE Transactions on Sustainable Energy*, 11. doi:10.1109/TSTE.2019.2926147.

Li, Z., Ye, L., Zhao, Y., Pei, M., Lu, P., Li, Y. et al. (2023) A spatiotemporal directed graph convolution network for ultra-short-term wind power prediction. *IEEE Transactions on Sustainable Energy*, 14. doi:10.1109/TSTE.2022.3198816.

Liao, K., Lu, D., Wang, M. & Yang, J. (2022) A low-pass virtual filter for output power smoothing of wind energy conversion systems. *IEEE Transactions on Industrial Electronics*, 69. doi:10.1109/TIE.2021.3139177.

Lin, W.H., Wang, P., Chao, K.M., Lin, H.C., Yang, Z.Y. & Lai, Y.H. (2021) Wind power forecasting with deep learning networks: Time-series forecasting†. *Applied Sciences (Switzerland)*, 11. doi:10.3390/app112110335.

Liu, Y., Guan, L., Hou, C., Han, H., Liu, Z., Sun, Y. et al. (2019) Wind power short-term prediction based on lstm and discrete wavelet transform. *Applied Sciences (Switzerland)*, 9. doi:10.3390/app9061108.

Olivares, K.G., Challu, C., Marcjasz, G., Weron, R. & Dubrawski, A. (2023) *Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx.*

Shen, X., Qu, Y., Huang, S., Li, Z. & Zhang, K. Wind speed data repairing method based on bidirectional prediction, *1 2021*. : Institute of Electrical and Electronics Engineers Inc., pp. 715–721.

Sun, Y., Tang, X., Sun, X., Jia, D., Cao, Z., Pan, J. et al. (2019) Model predictive control and improved low-pass filtering strategies based on wind power fluctuation mitigation. *Journal of Modern Power Systems and Clean Energy*, 7. doi:10.1007/s40565-018-0474-5.

Reports, 13, 1798. doi:10.1038/s41598-023-28390-w. URL https://doi.org/10.1038/s41598-023-28390-w

Wan, C., Xu, Z., Pinson, P., Dong, Z.Y. & Wong, K.P. (2014) Optimal prediction intervals of wind power generation. *IEEE Transactions on Power Systems*, 29. doi:10.1109/TPWRS.2013.2288100.

Wu, H., Meng, K., Fan, D., Zhang, Z. & Liu, Q. (2022) Multistep short-term wind speed forecasting using transformer. *Energy*, 261. doi:10.1016/j.energy.2022.125231.