VMD-Transformer-Based Wind Speed Forecasting for Spot Market Trading

Tian Dong[1], Yunfei Chu[2], Bicheng Tan[3], Yong Wang[2], Boqiang Li[2]
[1]Department of State Grid Jilin Electric Power Co. Ltd., Changchun 130022, Jilin, China
[2]Department of Economic and Technical Research Institute, State Grid Jilin Electric Power Co., Ltd., Changchun 130022, Jilin, China
[3]Department of State Grid Jilin Electric Power Co., Ltd. Jilin City Power Supply Company, Jilin 132000, Jilin, China

Corresponding author: Tian Dong (e-mail: dongtjl@163.com).

ABSTRACT In the power spot market, renewable energy power generation usually show very big fluctuation, so accurate wind speed forecast becomes more and more important. To solve this problem, this study propose a mixed forecasting framework which combine Variational Mode Decomposition (VMD) with a Transformer-based predict network: first, using VMD to decompose the complex wind speed series into many components of different time-scale, so that can dig out more richer and more structural dynamic feature; then, these refine signal will be put into Transformer model, who is good at catching long-short term time dependency, in order to exactly extract the micro hidden pattern in wind speed fluctuation. The experiment results show that the VMD–Transformer framework has quite strong generalization ability and also high prediction accuracy, so overall it works pretty nicely. The proposed method provides a reliable forecasting tool for renewable energy operation. It also gives effective decision support for power dispatching and participation in the electricity spot market.

INDEX TERMS variational mode decomposition; Transformer; wind speed forecasting; electricity spot markets

I. INTRODUCTION

WITH the gradual promotion of the electricity market reform, China's electricity spot market is now entering a stage of steady expansion. At the same time, the installed capacity of renewable energy keeps increasing, and it has become an unavoidable trend for clean energy to participate in spot trading [1]. It is predicted that by 2030, clean energy will achieve full participation through market mechanisms. As an important support of renewable energy, wind power is expected to play a more and more important role in spot trading, and it may gradually replace traditional energy sources. Meanwhile, with the continuous development of electromagnetic simulation, power system analysis, and other computationally intensive technologies, the demand for large-scale numerical calculations has been continuously increasing in modern energy systems. These tasks usually require substantial computational resources and electricity consumption, which puts forward higher requirements for efficient computing strategies and accurate forecasting technologies to support the reliable operation of power systems. However, wind power has obvious fluctuation, intermittency, and limited dispatchability [2]. Also, there are differences between the forecasted output and the actual output. These factors bring great operating pressure to market dispatching. Since point forecasting is the basis for trading participation decisions, deviation settlement, and profit allocation, its accuracy is directly related to economic benefits and the long-term feasibility of renewable energy integration into the market. Therefore, improving the accuracy of renewable energy forecasting is not only essential for reducing operational uncertainty but also beneficial for supporting efficient computational and operational management in complex power and electromagnetic-related engineering applications. Under the power spot market mechanism, renewable energy generation usually shows even more fierce fluctuation. Existing statistic data shows that only about 55% wind power project's point forecast accuracy can over 50% [3]. So, improve the reliability of wind speed forecasting has become more and more important, this have big meaning for upgrade the efficiency of grid dispatching and lower the operating risk that power generation enterprise is facing. Therefore, improving the reliability of wind speed forecasting has become increasingly important, and it is of great significance for enhancing grid dispatching efficiency and reducing the operational risks faced by power generation enterprises.

In recent years, wind speed forecasting has received wide attention from researchers, this greatly push forward the development of many predict technology. Existing method can usually be divided into three categories: physical model, statistic method [4] and artificial intelligence (AI) driven algorithm [5]. In short-term forecast task, researcher usually rely on statistic modeling and data assimilation method, so that can fully use the high-frequency observation information. By contrast, medium-and-long-term forecast mainly rely on physical model to catch the dynamic feature of large-scale atmosphere circulation. With the rapid development of deep learning, statistic method is gradually merge into neural network architecture, so that can better represent the non-linear feature of wind speed. For example, literature [6] design a ultra-short-term forecast method, which combine dynamic time warping algorithm for similarity measure, fast correlation-based filter feature selection method, and LSTM network for time series modeling; literature [7] propose a medium-and-long-term forecast framework which integrate graph convolutional network, wind speed difference fitting and particle swarm optimization strategy; besides, literature [8] develop a mixed model that fuse Variational Mode Decomposition (VMD), whale optimization algorithm, LSTM network and sparrow search algorithm.

Although these methods have lack of interpretability somewhat, but they perform very excellent in mining the hidden pattern inside the data. Therefore, they can obtain higher forecast performance, and win wide recognition in this field. For example, literature [9] propose a mixed L-LG-S model which combine Support Vector Regression (SVR), LSTM and LightGBM, achieve high-accuracy ultra-short-term forecast; literature [10] use a multiple Artificial Neural Network (ANN) framework to do short-term forecast, which comprehensive consider wind speed, wind direction, active power and theoretical power curve; besides, literature [11] evaluate the performance of LSTNet, TPA-LSTM and DA-RNN in short-term forecast, the result shows that these three method all reach a quite high accuracy, but DA-RNN exist the problem of attention mechanism redundancy, meanwhile LSTNet and TPA-LSTM also need relative more computing resource.

In wind speed forecast, single AI method often faces the problem of forecast lag [12]. One effective way to solve this problem is decompose the original wind speed series into many intrinsic modes, model each mode separately, finally reconstruct each forecast result into the final forecast output. This method can successful separate the high-frequency fluctuation and low-frequency trend, so that effective ease the lag phenomenon, and upgrade the system's response ability. For example, [13] proposed a hybrid wind speed forecasting framework called VMD-Ts2Vec-SVR. This framework combines VMD, the time-series contextual representation method Ts2Vec, and Support Vector Regression (SVR). [14] introduced the DBO-VMD-TCN-Transformer model. In this model, the Dung Beetle Optimization (DBO) algorithm is used to optimize the VMD decomposition parameters of non-stationary wind speed series. After that, the processed components are fed into a Temporal Convolutional Network and a Transformer to achieve high-accuracy short-term forecasting. [15] proposed the TSO-VMD-BiLSTM method, which combines VMD-based feature extraction with the Tunicate Swarm Optimization algorithm for parameter tuning. The processed data are then input into a Bidirectional Long Short-Term Memory (BiLSTM) network, and this helps improve short-term forecasting performance. In a similar way, Ref. [16] proposed the WOA-VMD-CNN-Transformer model, where the Whale Optimization Algorithm (WOA) is used to optimize the VMD parameters.

To overcome the limitations of a single model in feature extraction, the inherent complexity of wind power fluctuations, and the challenges brought by parameter tuning, this study proposes a wind speed forecasting framework based on the VMD–Transformer architecture. VMD is used to decompose the historical wind speed series into many IMFs. By this way, the data structure can be simplified, meanwhile the key fluctuation pattern also can be separate out. These IMF components then combine with auxiliary feature matrix, to build a comprehensive factor matrix. After that, this factor matrix is put into Transformer network. The encoder does collaborative processing for each component and it relate feature, this helps the model to obtain a better global understanding for wind speed series, and can effectively catch the dependency relationship between different data source. So, the forecasting process becomes more robust and contains richer information. The main contributions of this study can be summed up as follows:

1) A VMD–Transformer based wind speed forecasting framework is proposed. By combining the multi-scale signal decomposition ability of VMD with the long-term dependency modeling ability of Transformer, this method makes the forecasting accuracy go up a lot.

2) A comprehensive input matrix made up of VMD decomposed components and auxiliary meteorological features is built, and used as the input for the Transformer network. This feature representation way fuse information from many data source, enhance the describe ability for the complex dynamic feature of wind speed.

3) Using real meteorology observation data to do plenty of experiment, and further carry out wind speed-power mapping analysis. The experiment results fully prove the effectiveness of propose framework, and its practical value in real engineering application.

II. METHOD

A. VMD METHOD

Compared with the traditional Empirical Mode Decomposition (EMD), VMD adopts a variational optimization framework. This significantly suppresses the problem of mode mixing, and therefore improves the accuracy and robustness of the decomposition results. In a sense, EMD sometimes lets different signal patterns sit at the same table and talk over each other, while VMD tries to give each pattern its own seat. The main goal of this technique is to extract several signal components, and each component has a well-localized spectral distribution. At the same time, its energy is mainly concentrated around a specific centre frequency. This decomposition process can be rigorously described by the following variational formulation [17]:

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

In the above formula, u_k represents the k-th Intrinsic Mode Function (IMF), while ω_k denotes its corresponding center frequency. The term $\delta(t)$ represents the Dirac δ-function, and $x(t)$ denotes the original signal to be analyzed. To solve this constrained optimization problem, this study adopts the Augmented Lagrangian framework. By bringing in Lagrange multipliers, the original problem gets rebuilt into an unconstrained variational optimization problem, which makes it easier to solve. So, the optimization process can deal with the decomposition task in a more efficient and stable way, instead of struggling hard with the constraint conditions at every single step. The rebuilt problem can be shown as follows [18]:

$$\begin{aligned} L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \right. \\ \left. \cdot e^{-j\omega_k t} \right\|_2^2 \\ + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 \\ + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{2}$$

Here, α is the quadratic penalty factor, used to hold back the influence of Gaussian noise, so that the decomposition process won't get bothered easily by useless fluctuations. And $\lambda(t) \in \mathbb{R}$ stands for the Lagrange multiplier related to the constraint condition, it makes sure the optimization process follows the rules that are needed.

During this solution process, the variables u_k^{n+1}, ω_k^{n+1}, and λ^{n+1} are updated sequentially in an iterative manner, gradually moving closer and closer to the optimal solution, almost step by step and piece by piece, until convergence is finally achieved:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) - \sum_{i>k} \hat{u}_i^{n}(\omega) + \hat{\lambda}^n(\omega)/2}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega} \tag{4}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{u}(\omega) - \sum_{k=1}^{K} \hat{u}_k^{n+1}(\omega) \right) \tag{5}$$

This iteration process will keep going step by step, until the pre-set convergence criteria get satisfied at last. Through this mechanism, the VMD algorithm divides the wind speed time series into multiple sub-signals, and each sub-signal exhibits its own unique spectral characteristics, with a frequency signature that is quite different from the others. This decomposition method can not only achieve multi-scale feature extraction, but also generate more refined input features, thereby improving the performance of the subsequent forecasting model [19]. It should be noted that VMD is performed offline before model training, and the IMFs obtained from the decomposition are used as fixed input features for the subsequent Transformer forecasting model.

B. TRANSFORMER MODEL

The Transformer framework was proposed by Vaswani et al. in 2017, and it quickly attracted attention because it looked at sequence modeling from a rather different angle. This framework abandons the traditional recurrent structure of RNNs. Instead, by introducing the self-attention mechanism, it achieves parallel processing of sequential data, which greatly improves computational efficiency. It is almost like replacing a single-file queue with several open lanes, allowing information to move much more freely. Its architecture is shown in Fig. 1. For single-series time-series forecasting tasks, such as wind power or photovoltaic power forecasting, only the encoder is usually employed to extract significant features and generate high-accuracy forecasting results. Rather than carrying extra baggage, the model keeps the part that is most useful for the job and focuses its attention where it matters most.

The running process of this mechanism can be split, kind of naturally, into three stages. First, the input representation is turned into the query vector Q, key vector K, and value vector V through linear projection. Then, the alignment level between the query vector and the key vector is calculated to work out the attention scores. After that, the SoftMax function is used to normalize these scores, and in this way, the final attention weights are obtained:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} \right) V \tag{6}$$

The final output of the self-attention module is obtained by a weighted combination of the attention weights and the corresponding value vector V. To let the model catch more different patterns, in a way that feels a bit wider and richer, the Transformer uses a Multi-Head Attention strategy. That means several self-attention operations are carried out in parallel at the same time, and then their outputs are joined together. With this kind of structure, the representation ability of the model is not only expanded, but

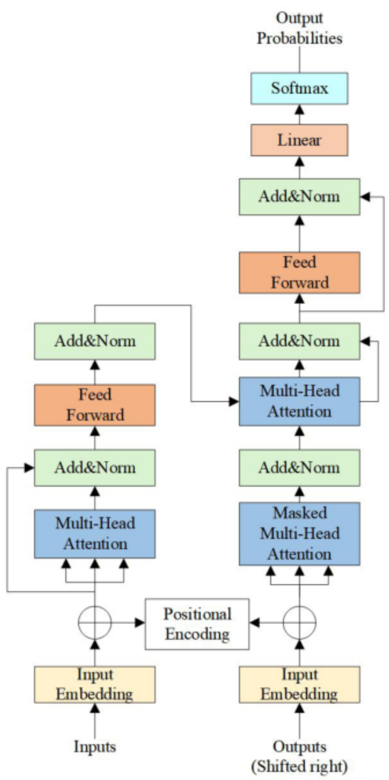

FIGURE 1. Schematic diagram of the Transformer model architecture.

its power to model the complex dependency relationships hidden inside the data is also improved quite a lot.

Besides, the Transformer bring in a position encoding strategy, it uses sine and cosine function to describe the relative position inside the sequence. These position encodings are combined with the input Embeddings, so it can make up for the inherent weakness that self-attention mechanism itself isn't sensitive to time order. Each Encoder sub-layer first has a Multi-Head Attention module, then followed by residual connection and Layer Normalization, kind of a fixed combo. All these designs work together to help the model converge more steadily and also make the training process much more efficient.

C. DESIGN OF THE VMD-TRANSFORMER FORECASTING MODEL

The integration of VMD and the Transformer framework combines multi-scale signal decomposition with an advanced ability to model dependency relationships. As shown in Fig. 2, this method can well deal with several key challenge in wind speed forecast, such as complex signal pattern and long-distance time correlation. Compare with existing VMD-LSTM and VMD-TCN method, the VMD-Transformer framework proposes in this study show stronger ability when modeling long-term time dependency. Although LSTM can catch sequence information quite well, but when dealing with long input sequence, its performance may decline somewhat because of information decay and limit parallel computing efficiency. TCN (Temporal Convolutional Network) enlarge the receptive field through dilated convolution, however, its ability to catch dependency relationship still be limited by its inherent convolution structure. By contrast, Transformer use self-attention mechanism, this makes it can directly build connection between all-time position in the sequence. Because of that, it can extract long-range dynamic patterns and complex nonlinear dependencies from wind speed data in a more efficient way. When combined with the multi-scale features produced by VMD, the proposed framework can explore both local fluctuation characteristics and global temporal information at the same time, and this finally leads to a clear improvement in forecasting accuracy as well as generalization performance.

(1) The precise separation of multi-scale features, together with the Transformer's ability to model global dependency relationships, fits very well with the naturally multi-time-scale characteristics of wind speed signals. More specifically, the low-frequency components are used to capture long-term seasonal and interannual variations, while the middle-frequency components reflect daily cycles and meteorological patterns. The high-frequency components, on the other hand, stand for fast fluctuations such as gusts and turbulence, jumping around here and there quite quickly. Through a variational optimization process, VMD performs adaptive decomposition and separates these different fluctuation patterns into independent IMFs. In this way, it creates structured inputs for the downstream modeling stage. Inside each extracted IMF, the Transformer's self-attention mechanism, especially its Multi-Head Attention module, can model both long-term temporal dependencies and the interactions among different components at the same time. By doing so, it builds a comprehensive and multi-scale feature representation.

(2) Data dimension alignment and feature integration. The IMFs generated by VMD are first normalized, and then rebuilt into tensors that are compatible with the dimensions of the meteorological variables. After that, they are concatenated together to form a unified input matrix. This step makes sure that the decomposed wind speed signals and the external meteorological data stay consistent in both dimensions and statistical characteristics, which, in a rather neat way, creates a homogeneous input space suitable for the Transformer's attention operations. Inside the Transformer encoder, these multimodal inputs are projected into a high-dimensional Latent Space through Embedding layers combined with positional encoding.

(3) Model training and optimization. The proposed forecasting framework is made up of two stages that are connected one after another. First, VMD is used as a fixed preprocessing technique to decompose the original wind speed sequence into several IMFs. After that, these decomposed components are combined with auxiliary meteorological features to build the input matrix. Then, the generated features are fed into the Transformer network for model training and prediction. During the training process, only

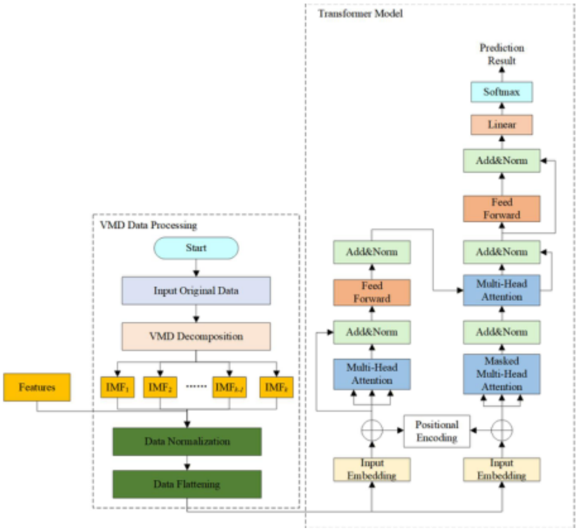

FIGURE 2. Architecture of the VMD-Transformer-based wind speed forecasting model.

the parameters of the Transformer are optimized, while the VMD decomposition part stays unchanged.

To sum up, the VMD-Transformer forecasting framework proposed in this paper achieves an effective fusion of multi-scale feature extraction through variational decomposition. The VMD module produces denoised and clearly structured multi-scale signal representations, and then the Transformer takes these representations and learns from them, capturing both global contextual information and long-range temporal correlations at the same time. Through the close combination of data preprocessing and feature representation, every part of this framework work together. What's more, this framework not only build a solid theoretical base for high-accuracy wind speed forecasting, but also gives practical support for renewable energy to connect into the grid safely and reliably.

The forecast goal of the framework propose in this paper is predict the wind speed of future 24 hour based on historical observation data. That is to say, this model not only forecast single time step forward, but try to describe the change trend of wind speed for a whole future day. Therefore, this task belongs to typical multi-step time series forecast problem, which means need to continuous predict a series of future value base on past observation, rather than a single independent value.

D. METHOD FOR WIND POWER ESTIMATION

The power output of wind power generation is decided together by the measure wind speed value and the specific operating characteristic of wind turbine. Then, the electric energy that generate will be convert into AC power through power electronic system, and finally transmit into power system. The performance of a wind turbine is constrained by three key wind speed thresholds, namely the cut-in wind speed v_{ci}, the rated wind speed v_r, and the cut-out wind speed v_{c0}. Because of these limits, the output power naturally shows a nonlinear behavior rather than changing in a simple straight-line way. As a result, the wind power output P_{WT} can be expressed differently across different wind speed ranges, almost like the turbine follows different rules in different wind conditions, as shown below:

$$P_{WT} = \begin{cases} 0, & 0 \le v \le v_{ci},\ v > v_{c0}, \\ P_r \dfrac{v^3 - v_{ci}^3}{v_r^3 - v_{ci}^3}, & v_{ci} \le v \le v_r, \\ P_r, & v_r \le v \le v_{c0}. \end{cases} \tag{7}$$

where, v represents the instantaneous wind speed, while P_r stands for the rated capacity of the wind turbine. The threshold parameters are all determined according to the design specifications of the turbine itself. When combined with the VMD-Transformer forecasting framework, the method first predicts the wind speed time series, and then these predicted values are fed into the wind power calculation model to generate the corresponding power forecasting results. In a way, the whole process moves in two linked steps: first the framework figures out how the wind is likely to behave, and then the power model quietly turns those wind speed predictions into estimates of future power output.

III. EXPERIMENTS

A. EXPERIMENTAL SETUP

The dataset used in this study contains 32,112 hourly meteorological records collected from January 1, 2022 to August 31, 2025. The data are divided into a training set and a testing set in an 8:2 ratio according to their chronological order. To build the forecasting samples, a sliding window strategy is adopted in this work. More specifically, the input sequence length is set to 48 time-steps, the forecasting horizon is set to 24 time-steps, and the sliding step size is 1. After the preprocessing stage, the training dataset contains 25,618 samples with a dimension of $25{,}618 \times 48 \times 15$, while the testing dataset contains 6,352 samples with a dimension of $6{,}352 \times 48 \times 15$. Since the DataLoader is configured with $drop_{last} = True$, which quietly discards incomplete batches at the edge, the final performance evaluation is carried out using 6,336 testing samples.

The input feature set contains a total of 15 variables, including 11 meteorological features and 4 mode components obtained through VMD decomposition. More specifically, the meteorological variables include air temperature, relative humidity, atmospheric pressure, precipitation, the zonal wind component, the meridional wind component, near-surface wind speed, wind direction, Global Horizontal Irradiance, Direct Normal Irradiance, and Diffuse Horizontal Irradiance. Besides these variables, VMD decomposition generates four Intrinsic Mode Functions, namely IMF1, IMF2, IMF3, and IMF4, and these components are incorporated into the input matrix to provide multi-scale temporal information. Under this configuration, meteorology feature describes the surrounding weather condition, while IMF component introduce detail information of different time-scale, this makes the model can observe wind speed data from many angles rather than single view, so that obtain a more comprehensive macro vision.

Transformer encoder is built by 4 EncoderLayers, and use 8 attention head. The dimension of Feed-Forward Network (FFN) is set to 512, Dropout rate is set to 0.5. For optimizer, the model use Adam optimizer, learning rate is set to 0.0003. Batch size is set to 64, the model total train 50 Epochs. Loss function use Mean Squared Error Loss (MSELoss), and random seed is fix at 100. Generally, speak, these setting build up the training strategy of the model, each parameter play its own role in keeping the learning process stable and guide the model to converge accurately.

B. VMD DECOMPOSITION OF WIND SPEED

The experiment dataset is composed by meteorology observation data collect from specific monitor station. The time range that data cover is from 00:00 Jan 1, 2022 to 23:00 Aug 31, 2025, in these 1338 days, total generate 32,112 records. In data pre-processing stage, use sliding window sampling method, finally build 31,970 effective sample. Among these sample, 25,618 is use for training, the rest 6,352 is leave for testing.

In the forecast process, the model input the accumulate meteorology data of past 48 hour, and base on this to generate the wind speed forecast of future 24 hour. Base on this setting, the forecast task discuss in this study can be view as a 24-step (24-hour) wind speed forecast problem. The complete input feature set total contains 15 independent variables, including the IMF component obtain through VMD (use for effective catch the multi-scale time pattern hidden in wind speed series), and another 11 meteorology and environment indicator. By fusing the decompose wind speed signal with this auxiliary parameter, the network can obtain a rich and multi-dimension representation about atmosphere condition. This makes the model can observe the change happen in environment more comprehensive, so that help to produce wind speed forecast result that not only more accurate, but also more robust when facing different weather condition.

All computing experiment is carried out on a Windows workstation which equip with Intel Core i5-10400 processor (main frequency 2.90 GHz) and 16 GB system memory. In a representative simulation, use VMD method to adaptive decompose the inherent non-stationary wind speed time series, transform it into a group of IMF component that cover the whole frequency range from low to high. This decomposes process catch the unique oscillation pattern across many time-scale, very helpful for understand the hidden characteristic of signal. As shown in Fig. 3(b)-3(e), the result clearly shows the step-by-step separate process of four frequency component from low to high.

To make sure the decomposition performance is stable, the VMD parameters are configured as following: the quadratic penalty factor α is set as 2000, the noise tolerance parameter τ is set as 0, the DC component option is disabled $(DC = 0)$, the center frequency is initialized as 1 uniformly $init = 1$, and the convergence tolerance is set as 1×10^{-7}. These settings help keep the decomposition process stable and well-performed. Among all the parameters, the number of decomposition modes K is a key factor that strongly affect the quality of signal decomposition. To find out the most suitable value, comparison experiments are done within the range of K from 3 to 7. The corresponding results are shown in Table 1, where the influence brought by different mode numbers can be compared side by side, kind of like lining them up to see which one win.

By analyzing the corresponding center frequency distribution, the spectral separation degree between different modes gets evaluated. The result show that, as K value increase, the center frequency of newly generated mode gradually gets close to the center frequency of existing mode. This show that the spectrum overlap phenomenon become more and more serious, and start to appear the potential risk of over-decomposition. After balancing the decompose ability and model complexity, finally choose $K = 4$ as the compromise plan. This number is neither too big nor too small, just stand at a quite suitable balance point. Under this setting, the center frequency of the four extracted modes is 0.000083, 0.013962, 0.040033, and 0.082768 separately. Based on the chosen decomposition layer number, the original wind speed sequence gets stripped into four Intrinsic Mode Functions (IMFs), as shown in Fig. 3. In some sense, the signal gets pulled apart layer by layer, so the different frequency characteristic hidden inside it can show up more clearly.

Fig. 3(b) shows the mode obtained through VMD decomposition that has the largest amplitude and the smoothest behavior. This mode represents the low-frequency trend background of the wind speed sequence. It reflects the quasi-stationary wind field formed jointly by large-scale climate evolution and local geographical conditions. Not only does it capture the long-term variation trend of wind speed, but it also provides a baseline background for wind speed forecasting. Compared with IMF1, Fig. 3(c) show fluctuation in higher frequency band, these fluctuations relate to weather-scale phenomena and the boundary layer daily cycle that get driven by solar radiation. This physical process can last from several hours to several days, and they are the main source of short-term wind speed fluctuation and daily periodic change. Because of this, they are one of the most important and relatively predictable feature components in short-term wind speed forecasting. Meanwhile, Fig. 3(d) describe the component with highest frequency and smallest amplitude. This component mainly come from atmospheric turbulence, intermittent gust, and measurement noise from monitoring equipment. Because of its essentially high non-linear and strong random nature, this component is usually regard as random disturbance. By effective separate this kind of component from the original signal, the algorithm provides a clearer input background for the model, make it can focus on the dominant physical process, so that more

TABLE I. Center Frequencies Obtained Under Different VMD Decomposition Levels

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
$K = 3$	0.000175	0.022888	0.076160				
$K = 4$	0.000083	0.013962	0.040033	0.082768			
$K = 5$	0.000081	0.013800	0.039587	0.080717	0.127018		
$K = 6$	0.000081	0.013747	0.039456	0.080271	0.121751	0.168289	
$K = 7$	0.000080	0.013728	0.039416	0.080184	0.120815	0.163971	0.212007

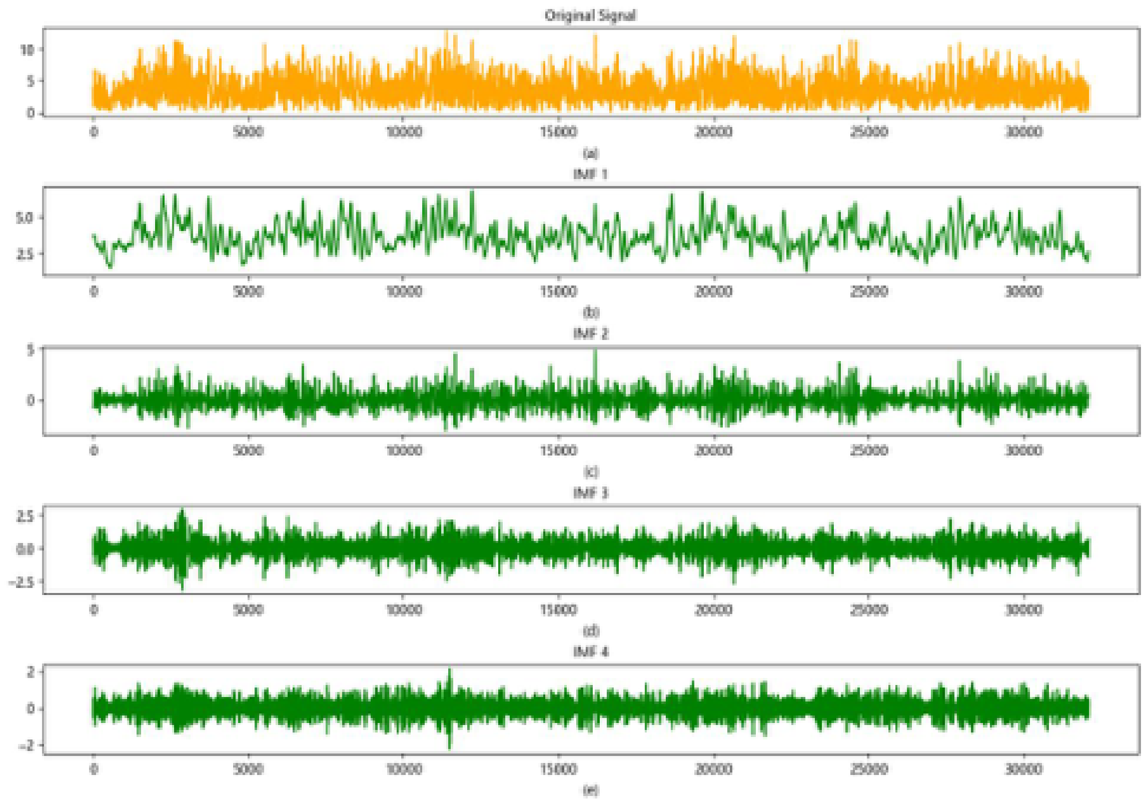

FIGURE 3. VMD decomposition of the raw wind speed data: (a) original data, and (b), (c), (d) and (e) represent the IMF1, IMF2, IMF3 and IMF4 components, respectively.

easily catch the truly meaningful dynamic feature hidden in wind speed data.

In summary, VMD successful decompose wind speed signal into trend component, periodic component and random component. By this way, it transforms a quite entangle and complex signal into many part that have more clear physical meaning. This decomposition builds a solid and physically-interpretable foundation for the forecasting task afterward, which let different physical process get modeled separately, instead of being mixed up together into one confusing whole lump.

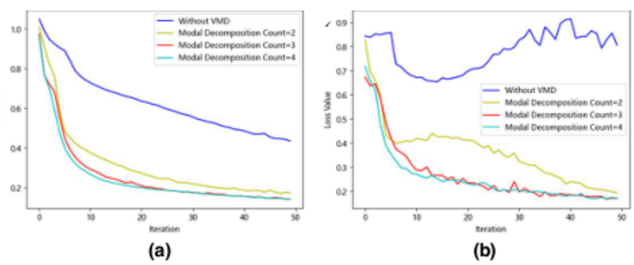

FIGURE 4. Loss value when the number of VMD modal decompositions is 2, 3, 4, and when no VMD decomposition is performed: (a) loss value of the training set, (b) loss value of the test set.

C. WIND SPEED FORECASTING

Several different scenarios were considered in the experiments, including using no VMD decomposition at all, decomposing the signal into 2, 3, or 4 components, and setting the time window length to 24 hours, 48 hours, or 72 hours. This allowed the effects of different decomposition levels and input lengths to be examined from more than one angle.

As shown in Fig. 4, when VMD decomposition was not applied, or when only 2 modal components were extracted,

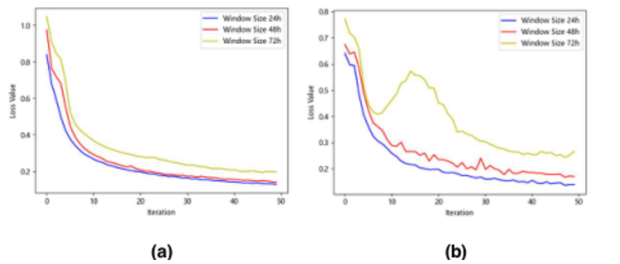

FIGURE 5. Loss value when the dataset time window is 24h, 48h, and 72h: (a) loss value of the training set, (b) loss value of the test set.

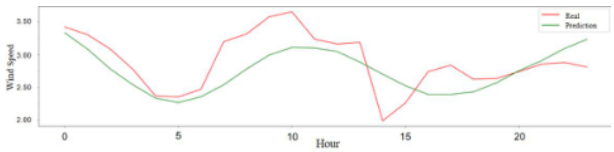

FIGURE 6. Comparison between predicted wind speeds and observed wind speeds in the early period.

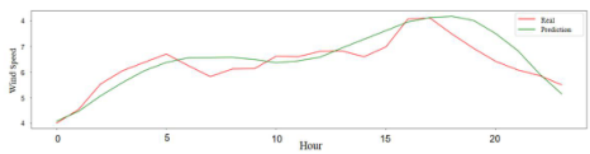

FIGURE 7. Comparison between predicted wind speeds and observed wind speeds in the middle period.

the model performed poorly on both the training set and the testing set. More specifically, in the case without VMD, the testing set failed to converge at all, which was quite a rough result. However, when the number of decomposition modes was set to 3 or 4, the results on the training and testing sets became much closer to each other. Based on these observations, the number of VMD modes was finally fixed at 4.

Fig. 5 further shows that the best performance on both the training set and the testing set was achieved when the time window length was 24 hours. By comparison, a 72-hour time window led to an extremely slow convergence speed on the testing set. The reason is fairly straightforward: when predicting wind speed for the next 24 hours, including data from the previous 48 or 72 hours introduces a good deal of redundant information. Compared with just using the data of recent 24 hours, this extra historical record might turn into an unnecessary burden, which add more load onto the training process, and at the end it brings negative effect to the model performance.

To fully evaluate the generalization performance of the VMD-Transformer model, this study randomly picks samples from the early, middle, and late stage of the test set to do forecasting analysis, as shown in Fig. 6–8. The result show that this model can keep giving stable and accurate forecasting throughout the whole evaluation period.

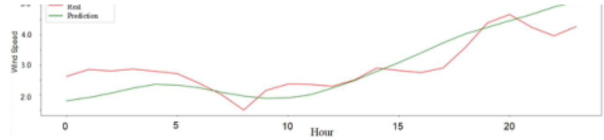

FIGURE 8. Comparison between predicted wind speeds and observed wind speeds in the late period.

To quantitatively evaluate the forecasting performance of the proposed model, this study use Mean Squared Error (MSE) as the main evaluation indicator. MSE is used to measure the average squared difference between forecasted wind speed value and actual observed value. Its definition is shown as the formula,

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Here, N represent the number of forecasting samples, y_i stand for the actual observed wind speed value, while $\hat{y}_i$ stand for the corresponding forecasted value. The smaller the MSE value is, the higher the forecasting accuracy of the model, simple logic really. In other words, lower MSE mean the forecasted wind speed sequence is closer to the actual observed sequence, and the gap between the two become less noticeable too.

For this 24-hour multi-step forecasting task, throughout the whole forecasting time horizon, the forecasted wind speed trajectory stays pretty close to the actual observed value. Among the samples randomly picked from early, middle, and late stage of the test set, their MSE (Mean Squared Error) value are 0.119, 0.259, and 0.277 separately. The experiment result show that, even at the middle and late stage of testing, this model still keeps an extremely high forecasting accuracy.

These result show that this model not only reach extremely high forecasting accuracy, but also show stronger temporal robustness. It can effectively adapt to the changing wind field condition, which help to make sure reliable and stable performance in real-world application. To further evaluate the forecasting ability in a quantitative way, this study calculates the average evaluation indicator for the three representative forecasting case shown in Fig. 6 to Fig. 8. The proposed model reaches an average coefficient of determination R^2 of 0.8069, MSE of 0.1638, Root Mean Squared Error (RMSE) of 0.4047, and Mean Absolute Error (MAE) of 0.3075. They show that, across different time period of the test set, the forecasted wind speed sequence keeps a high level of consistency with the actual observed data.

Besides forecasting accuracy, computing complexity and deployment cost are also key consideration factor in real-world application. In this study, the VMD-Transformer model gets trained for 50 epochs in total, and the whole training process only take 464 seconds, pretty fast if you ask me. This model contains 341,176 trainable parameters, and the saved model file size is around 1.33 MB. These result show that the proposed framework manages to keep computing and storage demand at a relatively low level, while still getting high forecasting accuracy at the same time. So, this model can be easily deployed on ordinary CPU-based or GPU-based computing environment, without needing to consume a huge amount of hardware resource. Because of this, it shows pretty good engineering practicality, and can provide useful support for real-world wind speed forecasting and power grid dispatching task. In some sense, it successfully manages to be both powerful and lightweight at the same time.

This excellent forecasting performance can be attribute to the complementary advantage between VMD and Transformer architecture, like two puzzle piece that just click together nicely. First, VMD decompose the original wind speed sequence into multiple mode component with different frequency characteristic, effectively separating long-term trend, periodic change, and random disturbance. This process largely reduces the non-stationarity and complexity of the original signal, making it easier for the network to learn its underlying evolving pattern. Second, Transformer use its Self-Attention mechanism, to model the important time dependency relationship both inside each decomposed component (intra-mode) and between different component (inter-mode). Because of this, it can efficiently extract the complicated time pattern and long-range dynamic information hidden inside the data. By combining multi-scale feature decomposition together with global dependency modeling, this framework can catch both local fluctuation characteristic and global time dynamic at the same time. This fusion of two-direction perspective gives the model a sharper understanding toward wind speed sequence, which then achieve a double improvement in both forecasting accuracy and generalization ability.

D. COMPARATIVE ANALYSIS OF WIND TURBINE POWER

In this study, the system performance is evaluated by using the forecasted wind speed value and the actual measured data. After that, according to the mathematical expression given in formula (7), the output power of wind turbine gets calculated from the wind speed value. In this calculation, standard wind turbine operating parameters are used: cut-in wind speed v_{ci} is set as 3 m/s, rated wind speed v_r is set as 11 m/s, cut-out wind speed v_{c0} is set as 25 m/s, rated power P_r is specified as 30 kW. Based on these setting, the forecasted wind turbine output power P_{WT} for the future 24 hours get generated. As shown in Fig. 9, the result clearly shows the time-varying change and amplitude size of the forecasted output power within the daily cycle.

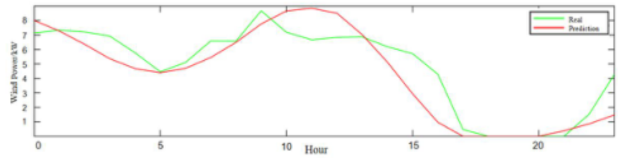

FIGURE 9. Comparison of wind turbine power generated by predicted wind speeds and observed wind speeds.

A pretty interesting point is that, between 18:00 and 20:00, the output power of wind turbine completely drops down to zero, which show that the turbine stops operating during this period. This mainly happen because the wind speed drop below the cut-in wind speed threshold that's needed to start and keep the turbine running. Especially encouraging is, during this period, the forecast power value and the actual observe power value reach zero at the same time. This provides extra evidence for the model's reliability, show that even in extreme situation when wind speed weaken and operating condition become more complex, this model still can accurately track the operating behavior of wind turbine, show very robust performance.

The power evaluates result report in this study, mainly aim to show the practical value of propose wind speed forecast framework in real wind power operating scenario. However, need to notice that, the decision in power spot market depends on many factors, rather than only rely on the accuracy of power forecast. Such as electricity price fluctuation, market regulation rule, unit commitment status and system dispatching constraint, also play important role. Therefore, the forecast output power should be regard as auxiliary support information for participate market and make operating plan, rather than directly use as optimization strategy for trading decision. Looking forward, future research will bring market mechanism and economic evaluation indicator into the analysis, so that the role of forecasting in electricity spot market operation can be explored in a deeper and more complete way.

IV. CONCLUSION

To meet the growing need of electricity spot market for high-accuracy wind speed forecasting, this study develops this model specially, to deal with the natural fluctuation and intermittency of renewable energy output. At the same time, it breaks through some common bottleneck that traditional method usually faces in feature extraction and forecasting performance. The experiment result show that this framework achieves stable convergence and stronger generalization ability. This can be seen from the steadily dropping MSE (Mean Squared Error) value on both training set and test set during the optimization process. Even when small fluctuation occasionally shows up, the model still keeps stable forecasting accuracy, without showing obvious performance degradation, pretty resilient little model honestly. By combining the multi-scale decomposition ability of VMD toward wind speed signal together with the ability of Transformer to catch long-range time dependency, the proposed method integrates the technical advantage of both. Through this way, it overcomes many common shortcomings found in single-model forecasting method, and bring a forecasting result that's improved a lot. Besides, the turbine-driven effect generated by the forecasted output power match highly with the actual measured value, which strongly prove the practical value of this model in real operating environment. Overall, this framework provides a reliable forecasting tool for renewable energy operation, and can give useful reference information for power dispatching and market participation. However, the economic benefit related to market trading decision haven't been fully explored yet at this point. Future research needs to combine market simulation and optimization method, to look into this part more deeply, so that the complete value of forecasting in market operation can be fully understood.

It needs to be pointed out that, the dataset used in this study is only collected from a single observation station. Because of this, this data mainly reflects the unique natural climate and geographic feature of that specific region. Although the experiment result show that the proposed model reaches a pretty satisfying forecasting accuracy and stability on the test dataset, its generalization ability still needs further verification under different climate condition, geographic location, and wind farm configuration. In other words, the model performs great here, but it still needs to prove itself in a wider environment, before a more universal conclusion can be made. Besides, this study mainly focuses on wind speed forecast and its follow-up power evaluate process. Several important factor, such as the uncertainty of Numerical Weather Prediction (NWP), the actual operating status of wind turbine and the price fluctuation of power market, have not been clearly consider in the model yet. Therefore, the propose framework is more suitable to be regard as a forecast auxiliary support tool for wind power operating and dispatching, rather than a complete solution that can deal with all forecast task and all market decision scenario.

Future research work will introduce multi-region dataset, heterogeneous meteorology information and power market operating mechanism into this framework. Through these works, researcher can more deeply discuss the robustness, expandability and practical application value of this method under more diverse condition. In the long run, with the continuous push of power market reform, and renewable energy generation expect to dominate the spot market trading by 2030, the VMD-Transformer method proposes in this study show a broad application prospect. It has potential to become an efficient forecast tool that support renewable energy operating and help future marketization application and the skyrocketing demand for electromagnetic simulation cases, and in these future scenario, accurate forecast will become more important than ever before.

ACKNOWLEDGMENT

This research was supported by 2025 High-Quality Development Strategy Research Project of the State Grid Jilin Electric Power Co., Ltd. (SGJLJY00ZLJS2500039).

REFERENCES

[1] Z. H. Huang, "Research on the trading mode of electricity spot market under the high proportion of new energy," *Market Weekly*, 2025.
[2] C. Chen, S. J. Yuan, Z. L. Yin, *et al.*, "A quantitative evaluation method for the volatility of distributed generation power time series," *J. Electron. Inf. Technol.*, 2022.
[3] Q. Wu and X. Zhu, "Optimal bidding strategy of a wind power producer in Chinese spot market considering green certificate trading," *Environmental Science & Pollution Research*, vol. 31, no. 9, 2024.
[4] X. L. Liu, Z. Lin, and Z. M. Feng, "Short-term offshore wind speed forecast by seasonal ARIMA—A comparison against GRU and LSTM," *Energy*, vol. 227, Art. no. 120492, 2021.
[5] K. Y. Shi, D. X. Zhang, X. Q. Han, *et al.*, "Digital twin model of photovoltaic power generation based on LSTM and transfer learning," *Power Syst. Technol.*, vol. 46, no. 4, pp. 1363–1372, 2022.
[6] Z. Q. Dong, "Ultra-short term wind speed prediction based on DTW-FCBF-LSTM model," *Electr. Meas. Instrum.*, vol. 57, no. 4, pp. 93–98, 2020.
[7] Z. H. Chen, W. Teng, X. F. Xu, *et al.*, "Long-term wind power prediction based on graph convolutional network and wind speed difference fitting," *China Electr. Power*, vol. 56, no. 10, pp. 96–105, 2023.
[8] R. Hu, J. F. Qiao, Y. H. Li, *et al.*, "Long-term wind power prediction based on WOA-VMD-SSA-LSTM," *Sol. Energy*, vol. 45, no. 9, pp. 549–556, 2024.
[9] M. Ashraf, B. Raza, M. Arshad, *et al.*, "Performance enhancement of short-term wind speed forecasting model using realtime data," *PLoS One*, vol. 19, no. 5, p. 19, 2024.
[10] R. K. Reja, R. Amin, Z. Tasneem, *et al.*, "A new ANN technique for short-term wind speed prediction based on SCADA system data in Turkey," *Atmosphere*, vol. 14, no. 10, Art. no. 1516, 2023.
[11] B. Huang, Y. Liang, and X. Qiu, "Wind power forecasting using attention-based recurrent neural networks: A comparative study," *IEEE Access*, vol. 9, pp. 40432–40444, 2021.
[12] S. Wang, C. Liu, K. Liang, *et al.*, "Wind speed prediction model based on improved VMD and sudden change of wind speed," *Sustainability*, vol. 14, no. 14, Art. no. 8705, 2022.
[13] S. Parri, K. Teeparthi, and V. Kosana, "A hybrid VMD based contextual feature representation approach for wind speed forecasting," *Renew. Energy*, vol. 219, Art. no. 119391, 2023.
[14] S. Zhang, C. Zhu, and X. Guo, "Wind-speed multi-step forecasting based on variational mode decomposition, temporal convolutional network, and transformer model," *Energies*, vol. 17, no. 9, Art. no. 1996, 2024.
[15] Q. Wang and L. Zhang, "Short-term wind speed prediction of wind farm based on TSO-VMD-BiLSTM," *PeerJ Computer Science*, vol. 10, 2024.
[16] Y. Liu, S. C. Wang, H. Liu, *et al.*, "Ultra-short-term wind speed combined prediction based on WOA-VMD-CNN-Transformer," *Comput. Syst. Appl.*, vol. 33, pp. 1–11, 2024.
[17] S. Li, L. Guo, J. Zhu, *et al.*, "Medium-term offshore wind speed multi-step forecasting based on VMD and GRU-MATNet model," *Ocean Eng.*, vol. 325, Art. no. 120737, 2025.
[18] M. Zhao, G. Guo, L. Fan, *et al.*, "Author correction: Short-term natural gas load forecasting based on EL-VMD-Transformer-ResLSTM," *Sci. Rep.*, vol. 14, no. 1, 2024.