Electrical Engineering (2025) 107:2903–2933
https://doi.org/10.1007/s00202-024-02685-1

ORIGINAL PAPER

Short-term wind power forecasting using the hybrid model of multivariate variational mode decomposition (MVMD) and long short-term memory (LSTM) neural networks

Ehsan Ghanbari[1] · **Ali Avar**[2]

Received: 6 April 2024 / Accepted: 13 August 2024 / Published online: 26 August 2024

Abstract
This paper presents a novel hybrid forecasting procedure for wind power using meteorological and historical data. The introduced method consists of three parts: effective feature selection, time series decomposition, and forecasting each decomposed time series. The minimum redundancy and maximum relevance (mRMR) algorithm is first utilized to choose the most effective features. In this stage, those selected historical features whose values are needed at the prediction time will be decomposed by the variational mode decomposition (VMD) technique and then forecasted by the long short-term memory (LSTM) networks. Then, the multivariate variational mode decomposition (MVMD) algorithm is exploited to simultaneously decompose selected features to address frequency mismatches between different series and capture the correlation among them. Given that various series and variables are involved in wind power forecasting, considering the correlation among them significantly affects prediction results. Afterward, LSTM neural networks are utilized to forecast each decomposed time series. Finally, two cases and several evaluation criteria are elaborated to assess the performance of the presented method. Experimental results confirm that the developed hybrid model, compared to the VMD-LSTM model, results in a decrease of 9.97, 4.33, and 3.32% in root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), respectively. The mean values of these criteria are, respectively, 4.6, 3.5, and 20.8 for the proposed model.

Keywords Wind power forecast · Multivariate variational mode decomposition · Long short-term memory · Minimum redundancy and maximum relevance technique · Prediction error

1 Introduction

During recent decades, the increasing environmental pollution and ever-rising growth of energy demand have made the utilization and development of renewable energy an absolute necessity. Fossil fuel reliance exacerbates climate change, contributing to environmental degradation and jeopardizing ecosystems. Transitioning to renewable energy sources is crucial for reducing greenhouse gas emissions, mitigating climate change, and safeguarding the planet's health. Beyond environmental concerns, renewable energy enhances energy security, as it offers a sustainable and diverse alternative to finite fossil fuel reserves. Additionally, the renewable energy sector fosters economic growth, job creation, and technological innovation, providing a pathway toward a resilient and sustainable energy future. Embracing renewable energy is not just an environmental imperative; it is an essential step toward building a more secure, resilient, and economically vibrant global society. According to studies carried out, wind power is the most cost-effective renewable energy [1]. The advantages of wind power are not limited to the aforementioned issues; wind energy resources have the technical capabilities to promptly react to market signals, thus providing a wide range of grid services in capacity, energy, and ancillary service markets [2, 3]. However, despite all of its environmental and economic benefits, wind energy is highly intermittent and volatile, which causes it to give rise to some major challenges and problems in the operation and management of the power system. As an effective avenue, enhancing the precision of wind power prediction can solve these problems, thereby increasing the utilization

✉ Ali Avar
a.avar1994@gmail.com; a.avar@modares.ac.ir

[1] Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran

[2] Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

rate of wind energy and making it develop faster. In this regard, it is crucially significant to forecast wind energy in an accurate way [4]. Recently, there has been a proliferation of methods put forth to enhance the accuracy of wind power forecasts. These methods can be divided into four main groups: physical methods, statistical approaches, artificial intelligence approaches, and hybrid approaches. Physical approaches construct a specific scenario based upon various meteorological parameters such as temperature, air pressure, atmospheric density, and humidity so as to turn weather prediction into wind power forecasting [5, 6]. In physical methods, the numerical weather prediction (NWP) model has been extensively used for the future wind power inference [7]. In Ref. [8], a Kalman filter is employed to generate the NWP model output to forecast wind speed. De Giorgi et al. [9] has introduced a hybrid wind power prediction procedure based upon the NWP model and artificial neural networks (ANNs) to improve the performance in long time horizons. Although physical methods do not need the input of historical information, they require extensive physical data, making the precision forecasting of wind power a challenging task [10]. Statistical methods utilize historical data to construct a model revealing the corresponding relationship between the original information and the actual forecasted outcome. In comparison with physical methods, statistical methods have the feature of simpler calculation and do not require large-scale monitoring equipment, resulting in cost reduction. Typical statistical forecasting methods include time series methods [11, 12]; artificial neural network-based methods, such as autoregressive [13], autoregressive moving average [14], autoregressive integrated moving average [15], and fractional autoregressive integrated moving average [16, 17]; the Gaussian process [18]; the Kalman filter [19]; the similar day method [20]; the Bayesian method [21, 22]; and so on. However, the complex characteristics of wind power limit the learning ability of statistical methods for long-term time series. Indeed, the longer the prediction time horizon, the greater the prediction errors. During the last decades, with the ongoing advancement of artificial intelligence, methods based upon artificial intelligence have gained significant usage in wind power forecasting because of their remarkable self-organization, self-learning, and nonlinear mapping abilities [23]. The existing popular and commonly employed artificial intelligence methods include the extreme learning machine (ELM) [24], support vector machine (SVM) [25], back-propagation neural network (BPNN) [26], multilayer perception (MLP) [27], and LSTM [28]. For instance, [29] has developed a modified empirical mode decomposition (EMD) model so as to address the uncertainty associated with the number of ANN model inputs, thus enhancing the forecasting precision. In Ref. [30], a newly developed method has been compared with different single models, such as quantile regression neural network (QRNN), support vector quantile regression (SVQR), and so on, in which the results show that SVQR generates the longest training time. Reviewing these papers shows that the main drawbacks of ANN and SVM are the lack of a unified approach for hyper-parameter tuning and the time-consuming nature of processing wind power information, respectively. The remarkable characteristics of ELM, in comparison with ANN and SVM, are fast learning speed, good generalization performance, and uncomplicated structure [31], while the unpredictably produced internal network parameters give rise to random forecasting results. Indicating that the most significant unsolved problem in ELM practical application is how to acquire the most suitable architecture of the network, Martínez-Martínez et al. [32] have presented a new algorithm for an automatic selection of ELM network architecture, thereby obtaining an appropriate number of the hidden nodes in the network architecture. Quite a few articles have shown that single models are not able to accommodate the high volatility created by continual changes in wind power, so it is not proper to utilize a single model to forecast wind power. As a result, the limited forecasting performance of single models has led researchers to focus on hybrid models. The hybrid forecasting methods make the most of the benefits of each method, enhance the forecasting ability, and make model selection simpler. The hybrid forecasting models can be categorized into the following types: (a) Linear and nonlinear hybrid methods. The linear methods can capture the linear components of the data, and the nonlinear methods are utilized to model the nonlinear parts [33, 34]. To acquire the final predicted values, the predicted results of the two methods are superimposed. (b) Weight-based hybrid methods. Various methods are employed so as to predict wind power at the same time. Each method is of a weighting coefficient that is directly proportional to the forecasting performance it offers. The ultimate forecasting result is acquired through a weighted amalgamation of the forecasted results from each method [35, 36]. (c) The combination of data decomposition techniques and forecasting methods. Signal decomposition technologies are employed to decompose the complicated wind power series into simple subsequences. Then, each subsequence is modeled and forecasted. Afterward, the forecasted values of different sub-models are added to obtain the ultimate wind power forecasting result. The decomposition algorithm plays an important role in preventing sensor failure and data missing [37]. As mentioned in Ref. [38], the last hybrid methods have better forecasting performance, as they are capable of effectively reducing the non-stationary and nonlinear features of the initial wind power series. The prevalent decomposition techniques include wavelet decomposition (WD) [39, 40], phase space decomposition [41, 42], empirical mode decomposition (EMD) [43, 44], ensemble EMD (EEMD) [45–47], complementary EEMD (CEEMD) [48], variational mode decomposition (VMD) [49–52], and

improved variational mode decomposition (IVMD) [53, 54]. Nevertheless, there is no common standard for selecting the wavelet basis function in WD, restricting its application [55]. In phase space decomposition technology, it is difficult to acquire the embedding dimension and delay time. The EMD method and its derivatives, such as the EEMD and CEEMD methods, are dependent on experience and lack a mathematical theory basis for wind power signal decomposition [52]. While performing effectively when applied to univariate analysis, the VMD method and all its improved versions are limited to univariate analysis, neglecting the mutual relationship between various time series and variables in both frequency and time domains [56]. On the other hand, when it comes to wind power generation, different variables are involved, suggesting that it is not sufficient to take into account simply the univariate historical wind power for wind power prediction. Furthermore, using VMD for the sequential decomposition of various variables causes computation costs to increase and frequency mismatch issues to raise [57]. To decompose related variables in a simultaneous manner, the MVMD method has been presented and employed for different prediction problems such as ocean wave energy forecasting [58], evapotranspiration prediction [59], forecasting of surface soil moisture [60], and carbon price prediction [61]. Nevertheless, there are only a few, if any, studies applying MVMD to carry out wind power prediction taking into account both various related meteorological series and wind power.

Another factor playing a pivotal role in building reliable forecasting strategies when there are many candidate inputs, which is the case with wind power forecasting, is selecting proper input features. If the size of the input set is reduced as a result of feature selection, the prediction engines will be capable of better learning the input/output mapping function of the process, thereby enhancing forecast precision. Additionally, feature selection can result in faster computation speed and easier model diagnosis and interpretation. Various feature selection approaches for wind power prediction have been presented in the literature. For instance, in Ref. [62], the features have been extracted from wind power historical values. Xue et al. [63] utilize mutual information (MI) so as to extract the spatial correlation information among variables and target variables and then employs conditional kernel density estimation approaches for wind power prediction. Wang et al. [64] exploit wind direction and speed data in various NWP data to train the model. In general, there are three common feature selection methods, including filtering methods, packaging methods, and embedding methods. The precision of the packaging and embedding methods are higher than that of the filtering method, yet they are easy to cause over-fitting. On the other hand, the filtering method is less likely to give rise to over-fitting and can reduce the computation dimension [65]. The common filtering method is MI. It sorts the variables based on the standard of MI and then selects the top-ranked ones. This method can preprocess the variables before utilizing the forecasting model for prediction [66]. Nevertheless, MI simply takes into account the correlation between the variable and the target variable but neglects the redundancy, making the input dimension larger and reducing the computation efficiency of the model. Comparatively, the mRMR method based on MI considers correlation besides redundancy, making it able to not only reduce the input dimension but also enhance the computation efficiency. According to the extensive literature analysis [67], the critical variables for wind power forecasting are time, wind direction and speed, temperature, air pressure, and their historical values, in addition to those of wind power. It is usually observed that the initial delays of these factors have the greatest dependence on the objective vector [27]. As a result, the value of them at the prediction moment tends to be selected as the proper features for wind power prediction. However, the problem that emerges here is that although the future values of meteorological variables are usually available, those of wind speed and wind direction are not. Neglecting such important variables adversely affects the wind power prediction process, causing the forecasting accuracy to decrease.

Upon investigating the challenges of wind power forecasting methods, it is revealed that there are still some areas in research that require further investigation. Firstly, identifying the mutual connection between the wind power and corresponding meteorological variables in frequency and time domains is difficult. It is possible for the existing univariate mode decomposition approaches to give rise to frequency mismatches between various series, causing the loss of useful data among various variables. Secondly, wind power forecasting approaches usually consider only the past values of meteorological and historical variables and fail to take into account the value of variables at the prediction moment as a result of not having access to them, despite their great importance. Thus, in order to improve forecasting accuracy, it is critical to consider these values in power prediction models. A review of published papers on hybrid models for wind power forecasting is carried out in Table 1.

To fill the above-mentioned gaps, this study proposes a two-step novel MVMD-LSTM hybrid approach for short-term wind power prediction. In this procedure, at first, the MRMR feature selection method is used to select proper features. In this stage, those selected historical features whose values at the prediction time are needed will be decomposed by the VMD technique and then forecasted by LSTM networks. Then, the MVMD technique is employed to decompose all the input variables into multiple simple sequences. Afterward, the wind power is predicted utilizing LSTM neural networks. The primary contributions of the proposed article are outlined as follows:

Table 1 Detailed review of hybrid forecasting works in the literature

References	Forecasting method	Decomposition technique	Feature selection method	Mode alignment	Application
[4, 23]	Echo state network	VMD	–	×	Wind Power
[5]	Fuzzy cognitive map	IVMD	–	×	Wind power
[6]	SVR	CEEMD	–	×	Wind power
[7, 50]	ELM	VMD	–	×	Load +wind power
[9]	ANN	WT	–	×	Wind power
[10]	BLSTM	BLSTM	–	×	Wind power
[24]	Beetle antenna search	VMD	–	×	Wind speed
[28]	MLP	WT	–	×	Wind power
[29]	FNN	EMD	–	×	Wind speed
[30]	LSTM	EMD	–	×	Wind speed
[36, 51, 52]	LSTM	VMD	–	×	Wind power
[37]	SVR	IEWT	–	×	Wind pressure
[40]	SVR	WT	–	×	Wind speed
[44]	ELM	EMD	–	×	Wind speed
[46]	Random forest	EEMD	–	×	Wind speed
[47]	ELM	EEMD	–	×	Wind speed
[48]	ELM	CEEMD	–	×	Wind speed
[49]	BPNN	VMD	–	×	Wind power
[50]	ELM	VMD	–	×	Wind power
[53]	LSTM	IVMD	–	×	Wind power
[54]	Kernel ELM	VMD	–	×	Wind speed
[55]	Improved LSTM	VMD	–	×	Wind power
[56]	Random forest	MEMD	–	×	Solar radiation
[57]	SVR	MEMD	–	×	Peak load
[70]	LSTM	WT	MI	×	Electricity load + price
This paper	LSTM	MVMD	mRMR	✓	Wind power

1. Considering the inherent multivariate characteristics of wind power, a MVMD method is used to analyze multiple variables together, thereby improving predictability.
2. While having all the advantageous features of the VMD technique, such as adaptability, robustness, and the ability to address the non-linearity and non-stationarity of wind power, unlike it and its derivatives, the MVMD approach is able to capture the correlation between all the variables involved in wind power prediction, such as historical wind data and meteorological variables, leading to improved accuracy in wind power forecasting because the correlation between these variables significantly affects wind power generation.
3. In contrast with conventional univariate mode decomposition techniques, the MVMD technique enjoys the mode alignment characteristic, which involves the synchronization of modes with analogous frequency content across multiple variables. This property helps identify and separate various sources of variability that can affect wind power. Furthermore, the mode alignment property contributes to extracting features that capture the most significant patterns and trends in the data, thereby reducing the noise and enhancing the accuracy of wind power forecasting. Finally, after aligning the modes across various variables, it is easier to understand the relationship among different variables and the way they affect wind

power, which leads to making the results of the MVMD technique more interpretable.

4. A new robust hybrid forecasting method (MVMD-LSTM) is presented and used for short-term wind power prediction. Experiments and comparison analyses are performed so as to assess the efficiency of the introduced approach.

Figure 1 shows the proposed approach structure. As seen in this figure, this study begins with selecting the most optimal and relevant input feature set from the wind power and its influential factors using the mRMR feature selection method. Then, to comprehensively capture the correlations between the selected factors in the previous step in frequency and time domains and obtain synchronous time–frequency analyses of the factors, the MVMD algorithm is employed. Afterward, several sets of the sub-series obtained in the previous step are selected as the input data and used to conduct predictions for the next 24 h based on the LSTM network. Finally, the ultimate forecasting outcomes are acquired by summing the corresponding predicted values of each sub-signal.

The rest of this paper is structured as follows. Section 2 provides the basic methodologies of MRMR, MVMD, and LSTM, as well as the introduced hybrid modeling framework. The experimental research is presented in Sect. 3. The experimental results and analysis are reported in Sect. 4. Section 5 presents the discussion and summaries of the proposed modeling framework. Finally, the paper is concluded in Section 5.

2 Methodology

In this section, a complete description of the basic methodologies of the techniques employed in this paper and the proposed hybrid modeling framework are addressed.

2.1 Feature selection based on max-relevance and min-redundancy algorithm

When it comes to wind power forecasting, it is not proper to consider only one historical or meteorological variable as the relevant feature. This is because it is less likely for simply one feature to mirror the fluctuation features of the real wind power. In this regard, to enhance the precision of wind power forecasting, more features associated with fluctuation characteristics of wind power should be selected as input variables of deep neural networks. In addition, due to the fact that any algorithm's performance is closely linked to the data quality, it is pivotal to select the most representative subset of relevant features. In short-term wind power forecasting, some features may be redundant to the model and some may have no irrelevance to wind power, which adversely affects the performance of the algorithm. So, feature selected is needed to remove such features. Therefore, an appropriate set of features should have a strong correlation with wind power while also demonstrating minimal redundancy with other features. Widely employed metrics for correlation analysis include Spearman's and Pearson's correlation coefficients, yet the former is primarily employed to evaluate the linear relationship between two variables, and the latter is utilized so as to determine the monotonic relationship. However, in short-term wind power prediction, it is possible to exist a non-monotonic, nonlinear relationship between the variables and wind power. Under such conditions, the mutual information can be utilized to determine the relationship between the features and wind power. The mutual information between two variables is a measure of the degree to which knowledge of one variable reduces uncertainty about the other.

The mutual information I of the feature f_i and wind power P is determined as follows:

$$I(f_i, P) = \sum_{t,t' \epsilon T} p\left(f_i = f_i(t), P = P\left(t'\right)\right) \log \frac{p\left(f_i = f_i(t), P = P\left(t'\right)\right)}{p\left(f_i = f_i(t)\right) p\left(P = P\left(t'\right)\right)} \tag{1}$$

$$I\left(f_i, f_j\right) = \sum_{t,t' \epsilon T} p\left(f_i = f_i(t), f_j = f_j\left(t'\right)\right) \log \frac{p\left(f_i = f_i(t), f_j = f_j\left(t'\right)\right)}{p\left(f_i = f_i(t)\right) p\left(f_j = f_j\left(t'\right)\right)} \tag{2}$$

where p is the probability. $I(f_i, P)$ and $I(f_i, P)$ denote the mutual information between the ith feature and wind power, and that between the ith and jth features, respectively.

In this study, the relationship between wind power and historical and meteorological features is analyzed using a feature selection approach known as minimum redundancy and maximum relevance (mRMR). The reason behind using mRMR is to select the features that provide minimum correlation among themselves while having maximum correlation with the wind power. The relevance of a feature set (denoted as S) for the wind power (denoted as P) can be determined by calculating the average mutual information across all the individual features within the set (denoted as f_i) and P as follows.

The mRMR technique seeks to find an optimal set of S features maximizing V_S, that is, the relevance of S regarding an output variable P. It minimizes W_S, referring to the redundancy of S, where V_S and W_S are determined as follows:

$$V_s = \frac{1}{|S|} \sum_{f_i \in S} I(f_i, P) \tag{3}$$

$$W_s = \frac{1}{|S|^2} \sum_{f_i, f_i \in S} I(f_i, f_j) \tag{4}$$

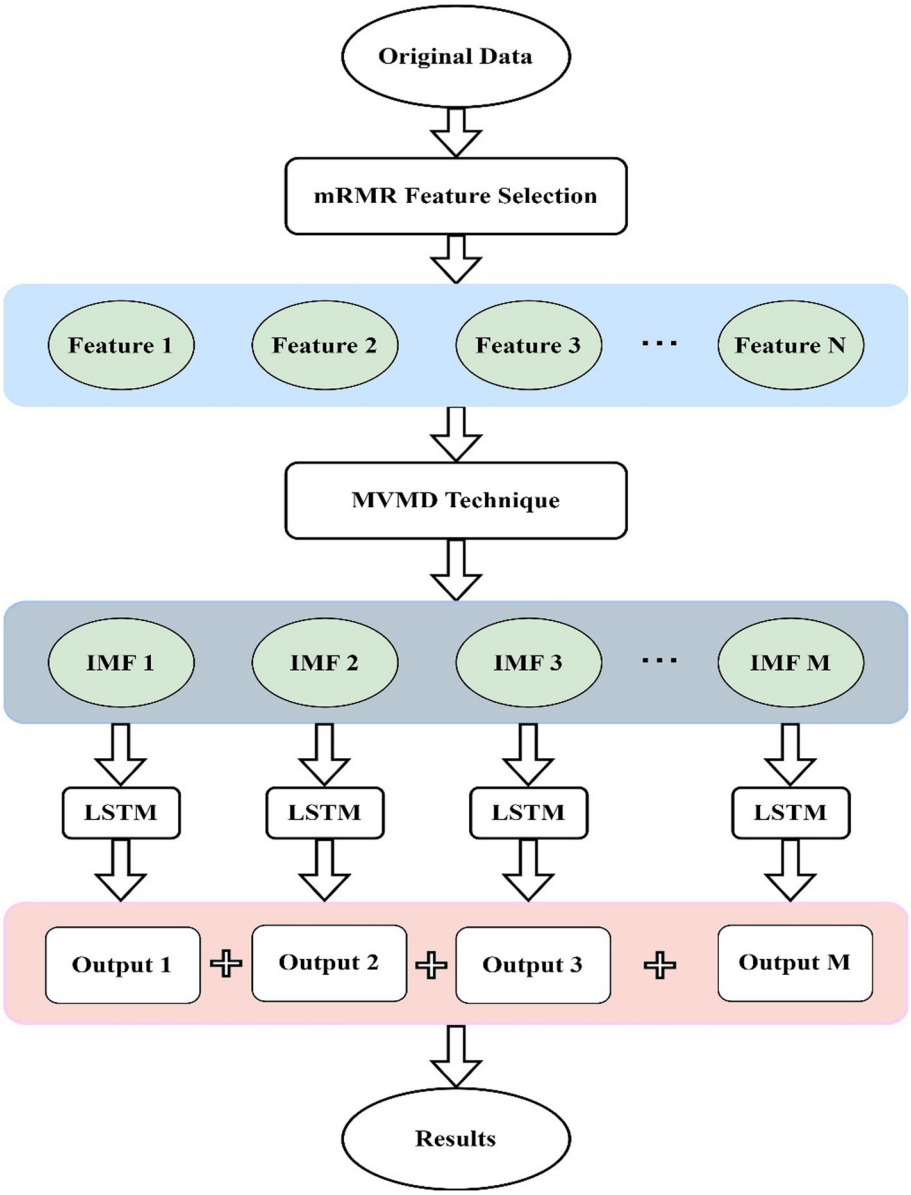

Fig. 1 Proposed approach structure

Here, $|S|$ is the number of features in S, a subset of historical and meteorological features.

The mRMR method sorts features as follows.

1. Selecting the feature having the highest relevance ($\max_{f_i \in \Omega_S} V_{f_i}$); adding it to the selected set S. ($V_{f_i} = I(f_i, P)$, and Ω_S is the whole feature set)
2. Searching for features relevant but not redundant with the target variable in S^c (the supplement of S), if none are found, proceed to step 4. Else, we should choose the most relevant feature ($\max_{f_i \in S^c, W_{f_i}=0} V_{f_i}$) and add the selected feature to S.
3. Repeating the previous step. If the redundancy of all features in S is not zero, then going to the next step.
4. In set S^c, selecting the feature which has the highest mutual information quotient (MIQ) value and has redundancy and relevance, and finally, adding it to S.

$$\max_{f_i \in S^c} \mathrm{MIQ}_{f_i} = \max_{f_i \in S^c} \frac{I(f_i, P)}{\frac{1}{|S|}\sum_{f_i \in S} I\left(f_i, f_j\right)} \tag{5}$$

5. Repeating the previous step, if the correlation of all features in S is zero, then go to the next step.
6. Adding the irrelevant features to S randomly.

2.2 Variational mode decomposition (VMD) algorithm

The VMD algorithm is described for univariate data in the following. The aim of VMD is to decompose an input data into K number of intrinsic principal modes u_k. Readers interested in a comprehensive discussion on the topic are referred to [68]. The mathematical framework of the variational problem can be expressed as

$$\min_{\{y_k\},\{\mathcal{W}_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) y_k(t) \right] e^{-j\mathcal{W}_k t} \right\|_2^2 \right\}$$
$$\text{Subject to} \quad \sum_k y_k = z(t) \tag{6}$$

where $\{y_k\} = \{y_1, y_2, \ldots, y_K\}$ and $\{\mathcal{W}_k\} = \{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_K\}$ denote the set of all modes and respective frequencies. In order to convert this constrained variational problem into an unconstrained optimization problem, the balancing parameter α and Lagrange multipliers λ are introduced to encourage reconstruction fidelity and enforce constraints strictly, as shown in the following equation

$$\mathcal{L}(\{y_k\}, \{\mathcal{W}_k\}, \lambda) = \alpha \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) y_k(t) \right] e^{-j\mathcal{W}_k t} \right\|_2^2 + \left\| z(t) - \sum_k y_k \right\|_2^2 \left\langle \lambda(t), z(t) - \sum_k y_k(t) \right\rangle \right\} \tag{7}$$

By utilizing the Alternate Direction Method of Multipliers (ADMM), different modes and their respective center frequencies would be regulated to acquire the optimal y_k, $\mathcal{W}_k$, and λ.

$$\hat{y}_k^{n+1} = \frac{\hat{z}(\mathcal{W}) - \sum_{i \neq k} \hat{y}_i(\mathcal{W}) + \frac{\hat{\lambda}(\mathcal{W})}{2}}{1 + 2\alpha(\mathcal{W} - \mathcal{W}_k)^2} \tag{8}$$

$$\hat{\mathrm{W}}_k^{n+1} = \frac{\int_0^\infty \mathcal{W} |\hat{y}_k(\mathcal{W})|^2 d\mathcal{W}}{\int_0^\infty |\hat{y}_k(\mathcal{W})|^2 d\mathcal{W}} \tag{9}$$

$$\hat{\lambda}^{n+1}(\mathcal{W}) = \hat{\lambda}^n(\mathcal{W}) + \tau \left(\hat{z}(\mathcal{W}) - \sum_k \hat{y}_k^{n+1} \right) \tag{10}$$

where $\hat{z}(\mathcal{W})$, $\hat{y}_i(\mathcal{W})$, and $\hat{\lambda}(\mathcal{W})$ denote the Fourier transforms of each variable and n is the number of iterations.

2.3 Multivariate variational mode decomposition (MVMD) algorithm

As the first fully multivariate extension of the VMD technique, the MVMD algorithm does not suffer from the constraints found in the current multivariate data-driven methods. For example, unlike the MEMD algorithm, MVMD does not encounter the problems of sensitivity to noise and sampling rate, existing mode mixing among its various channels, and not having a mathematical framework. In addition, in contrast with SST, MVMD has mode separation capability. Moreover, MVMD inherits quite a few of the useful characteristics of the standard VMD technique while being superior to it and even its channel-wise version in that it can recover multivariate modulated oscillations, or any useful mutual information associated with multiple channels, from input data, and that it has the mode alignment feature which involves matching or alignment modes that have analogous frequency content across various channels [68]. With this in mind, in this paper, the MVMD algorithm has been employed so as to conquer the inherent drawbacks of existing multivariate decomposition techniques and take advantage of the benefits provided by MVMD, thus making the forecasting accuracy more reliable. At first, it defines the multivariate mode, which can be represented in a vector form,

$$u_{k,n}(t) = \begin{bmatrix} u_{k,1} \\ u_{k,2} \\ \vdots \\ u_{k,N} \end{bmatrix} = \begin{bmatrix} a_{k,1} \\ a_{k,2} \\ \vdots \\ a_{k,N} \end{bmatrix} \cos(\varphi_k(t)) \tag{11}$$

Here N is the channel number; φ_i and a_i are the phase and amplitude function associated with the ith signal component, respectively. This definition takes into account the relationship between various channels of multivariate signals [69]. The original multivariate signal is $X(t) = [X_1(t), X_2(t), \ldots, X_N(t)]$, and

$$X(t) = \sum_{k=1}^{K} u_k(t) \tag{12}$$

where $u_k(t) = [u_1(t), u_2(t), ..., u_N(t)]$. Now, an analytical expression of vector signal $u_{k,n}(t)$ can be acquired by using the Hilbert transform operator,

$$u^+_{k,n}(t) = u_{k,n}(t) + j\mathcal{H}(u_{k,n}(t)) = \begin{bmatrix} a_{k,1} \\ a_{k,2} \\ \vdots \\ a_{k,N} \end{bmatrix} \exp(j\varphi_k(t)) \tag{13}$$

where $\mathcal{H}$ is Hilbert transform operator.

The aim of the MVMD technique is to extract a collection of modes $\{u_k(t)\}$, $k = 1, 2, \ldots, K$ from the input signal. These extracted modes fulfill two conditions: their combined bandwidth is minimized, and they can precisely reconstruct the original input data. We can determine the bandwidth of $u_k(t)$ by computing the L_2 norm of the gradient function of $u^+_k(t)$. Thus, the cost function of MVMD will be a multivariate extension of that employed in the corresponding optimization problem of VMD and is given by

$$f = \sum_k \left\| \partial_t \left[\mathrm{e}^{-j\omega_k t} u^+_k(t) \right] \right\|^2_2 \tag{14}$$

It should be noted that in the above equation, a single frequency component ω_k is utilized in the harmonic mixing of the entire vector $u^+_k(t)$. As a result, we look for multivariate oscillations in $u_k(t)$ that have the single shared frequency component ω_k among all the channels. So, the estimate of the bandwidth of modulated multivariate oscillations can be made through shifting the unilateral frequency spectrum of all the channels of $u^+_k(t)$ with ω_k and taking the Frobenius norm of the acquired matrix. The Frobenius norm has been adopted as an extension of the $L2$ norm, which is utilized in the original VMD algorithm to the matrices appearing as a result of multiple channel signal representation in MVMD. That results in this following simple representation for f

$$f = \sum_k \sum_n \left\| \partial_t \left[u^+_{k,n}(t) \mathrm{e}^{-j\omega_k t} \right] \right\|^2_2 \tag{15}$$

where $u^+_{k,n}(t)$ is the analytic modulated signal with mode number k and channel number n. Then, the constrained optimization problem of MVMD can be represented as follows.

$$\begin{aligned} &\min_{\{u_{k,n}\},\{\omega_k\}} \left\{ \sum_k \sum_n \left\| \partial_t \left[u^+_{k,n}(t) \mathrm{e}^{-j\omega_k t} \right] \right\|^2_2 \right\} \\ &\text{Subject to} X_N(t) = \sum_k u_{k,n}(t),\ n = 1, 2, \ldots, N \end{aligned} \tag{16}$$

It should be mentioned that in the above model, there are multiple linear equality constraints related to the whole number of channels. The related augmented Lagrangian function of this constrained optimization problem then can be given by

$$\begin{aligned} \mathcal{L}\left(\{u_{k,n}\}, \{\omega_k\}, \lambda_n\right) = \alpha \sum_k \sum_n \left\| \partial_t u^+_{k,n}(t) \mathrm{e}^{-j\omega_k t} \right\|^2_2 \\ + \sum_n \left\| x_n(t) - \sum_k u_{k,n}(t) \right\| \\ + \sum_n \left\langle \lambda_n(t),\ x_n(t) - \sum_k u_{k,n}(t) \right\rangle \end{aligned} \tag{17}$$

where α is the balancing parameter of the data-fidelity constraint;$\langle ., . \rangle$ and λ_n stand for the inner product and Lagrangian multipliers, respectively. Equation (12) is solved by using the ADMM method [68]. The ADMM approach turns the complicated optimization problem into several smaller suboptimization problems, which are simple to address. This is because these sub-problems involve minimizing the cost function by iterating over a single parameter of interest instead of minimizing it for all variables at the same time. The MVMD algorithm is demonstrated in Algorithm 1,

Algorithm 1 Demonstration of ADMM for MVMD Optimization

Initialize: $\left\{u_{k,n}^{1}\right\},\{\omega_{k}^{1}\},\lambda^{1},c\longleftarrow 0,\epsilon\longleftarrow 10^{-7}$

repeat

$c\longleftarrow c+1$

for $k=1:K$ do

for $n=1:N$ **do** *Update mode* $u_{k,n}$:

$$u_{k,n}^{c+1}\longleftarrow\arg\min_{u_{k,n}}\mathcal{L}(\left\{u_{i<k,n}^{c+1}\right\},\left\{u_{i\geq k,n}^{c}\right\},\{\omega_{i}^{c}\},\lambda_{n}^{c}) \tag{18}$$

end for

end for

for $k=1:K$ **do** *Update center frequency* ω_k

$$\omega_{k}^{c+1}\longleftarrow\arg\min_{\omega_{k}}\mathcal{L}(\left\{u_{i,n}^{c+1}\right\},\{u_{i<k}^{c+1}\},\{\omega_{i\geq k}^{c}\},\lambda_{n}^{c}) \tag{19}$$

end for

for $n=1:N$ **do** *Update* λ_n

$$\lambda_{n}^{c+1}=\lambda_{n}^{c}+\tau(x_{n}-\textstyle\sum_{k}u_{k,n}^{c+1}) \tag{20}$$

end for

until Convergence: $\sum_{k}\sum_{n}\frac{\left\|u_{k,n}^{c+1}-u_{k,n}^{c}\right\|_{2}^{2}}{\left\|u_{k,n}^{c}\right\|_{2}^{2}}<\epsilon$

where k, τ, and $\in$ are the predefined number of components, the Lagrange update parameter, and the convergence tolerance, respectively.

In this algorithm, the mode and center frequency are iteratively updated to obtain the best possible estimation while conforming to the constraints of the variational model. The mode update can be formulated as follows.

$$\hat{u}_{k\cdot n}^{c+1}=\frac{\hat{x}_{n}(\omega)-\sum_{i\neq k}\hat{u}_{i,n}(\omega)+\frac{\hat{\lambda}_{n}(\omega)}{2}}{1+2\alpha(\omega-\omega_{k})^{2}} \tag{21}$$

In addition, the center frequency can be updated by the following equation.

$$\omega_{k}^{c+1}=\frac{\sum_{n}\int_{0}^{\infty}\omega\left|\hat{u}_{k,n}(\omega)\right|^{2}\mathrm{d}\omega}{\sum_{n}\int_{0}^{\infty}\left|\hat{u}_{k,n}(\omega)\right|^{2}\mathrm{d}\omega} \tag{22}$$

where the hat symbol $(\hat{\cdot})$ denotes the Fourier transform and the parenthesis superscript indicates the iteration number (Fig. 2).

2.4 Long short-term memory (LSTM) neural networks

Recently, deep learning has created a new trend in machine learning. The deep learning basics have their roots in classic neural networks. However, they diverge from traditional neural networks because they are made up of several hidden layers of neural networks performing complex operations on massive amounts of structured and unstructured data. Hence, deep learning is a machine learning method using deep neural networks [70]. Dealing with time series forecasting issues, quite a few models have been introduced across various application domains. Recurrent neural networks (RNNs) have found successful applications in various machine learning problems. Such models are presented so as to deal with learning problems that are time dependent. Figure 3 shows the flow of information within the cells of the standard RNN. At every time step, h_t is the function of the current input x_t and previous cell state h_{t-1}.

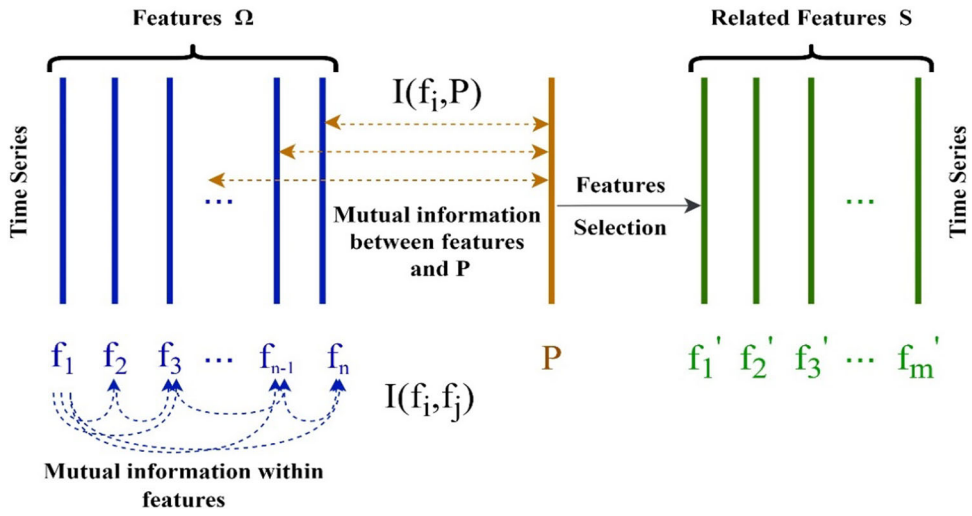

Fig. 2 Diagram of related factor selection considering mutual information criteria

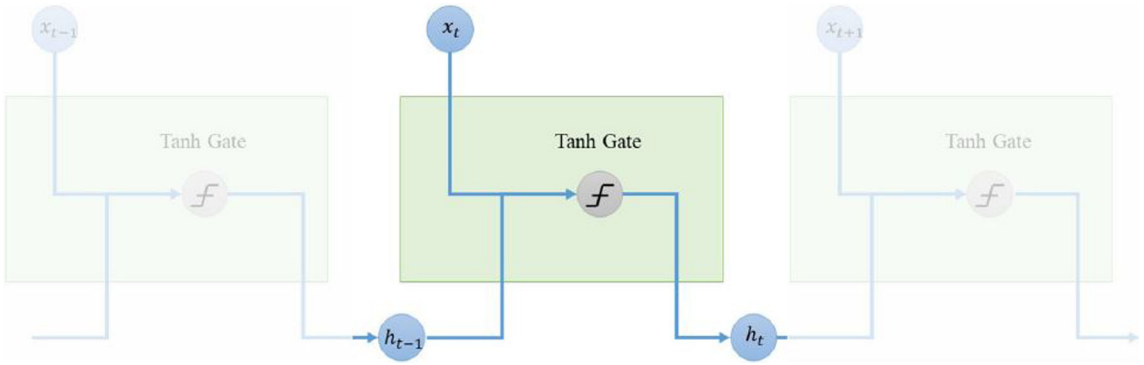

Fig. 3 Cell chain structure of recurrent neural networks (RNNs)

The back-propagation through time (BPTT) algorithm is used to train RNNs. This algorithm is utilized to update the network weights so as to minimize the error between forecasted and real outputs. This is achieved by propagating the error backward across the network and adjusting the weights accordingly. As for RNNs, BPTT goes forward through time, updates the cell state according to the previous state and input, generates an output of the time step, computes the loss at the very time step, and acquires the total loss through adding the losses from the individual time steps. In other words, RNN errors are back-propagated at every individual time step and then throughout the whole sequence of data. BPTT does matrix multiplication between every time step, and a nonlinear activation function is also employed to update the cells. As a result, numerous repeated weight matrix multiplications and utilizations of the derivative of the activation function are needed in order to compute the gradient. However, if the gradients or weight values are too small, we will encounter an issue referred to as the "vanishing gradient" problem. This occurs if the weight values progressively diminish in size due to performing the repeated multiplications, which in turn acts as a deterrent to the network training [71].

LSTM, as a special type of RNNs, is capable of learning short-term and long-term dependencies [72]. The LSTM units have memory cells linked across consecutive layers. Additionally, thanks to having various interacting layers, they are able to selectively regulate the information flow within units. The basic LSTM cell is shown in Fig. 4. The main component of a LSTM unit is a structure referred to as the "gate," enabling it to add information to or remove information from the cell states in a selective manner. The gates consist of a neural net layer such as a sigmoid or tanh as well as an element-wise multiplication. LSTM uses four basic steps to process the information:

- Forget: LSTM determines how much of the information is irrelevant and should be forgotten. It depends on previous internal stat h_{t-1} and input x_t, since part of that data might be insignificant.
- Store: LSTM determines how much of the information is significant and should be stored in its cell state to be used in next steps.
- Update: LSTM updates its cell state in a selective manner using the input and necessary part of the previous data.

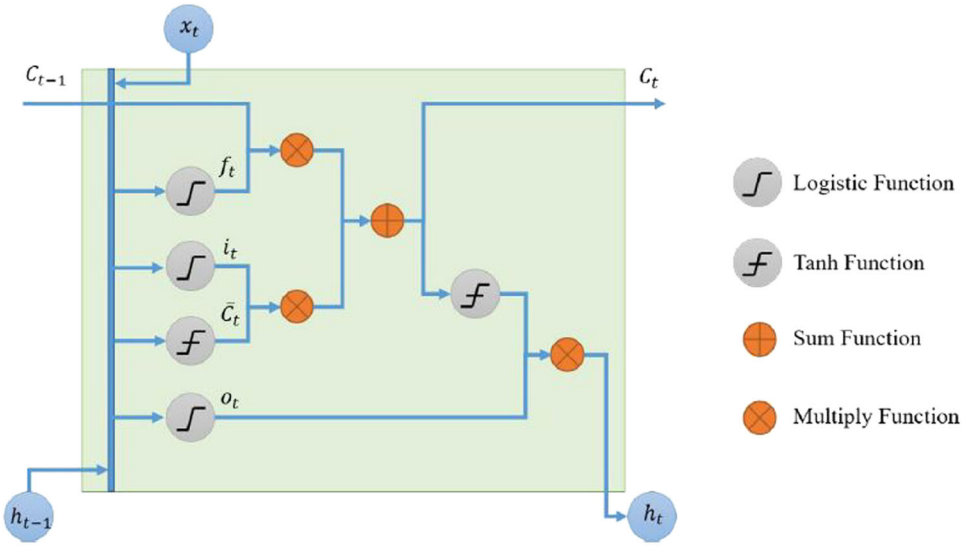

Fig. 4 Basic long short-term memory (LSTM) cell

- Output: LSTM decides what data encoded in the cell state should be passed to the network as the input in the next time step.

A significant characteristic of the LSTM unit is that it uses all various update mechanisms and gating to generate the internal cell state C, allowing for the continuous gradient flow across time. It acts as a type of cell state highway that alleviates and mitigates the vanishing gradient problem. The cell state of the LSTM unit is able to accumulate various activities over time. Due to the fact that derivatives distribute across sums, the error derivatives no longer quickly vanish while propagated backward in time. As a result, the LSTM unit is able to handle tasks across long sequences and discover features extended over a long range [73]. Definition of activation:

$$\text{sigmoid}(z) = \frac{1}{1+\mathrm{e}^{z}} \tag{23}$$

$$\tanh(z) = \frac{\mathrm{e}^{z}-\mathrm{e}^{-z}}{\mathrm{e}^{z}+\mathrm{e}^{-z}} \tag{24}$$

The LSTM information flow is formulated as follows.

$$f_t = \text{sigmoid}\left(W_f\left[x_t, h_{t-1}\right] + b_f\right) \tag{25}$$

$$i_t = \text{sigmoid}\left(W_f\left[x_t, h_{t-1}\right] + b_i\right) \tag{26}$$

$$\overline{C}_t = \tanh\left(W_f\left[x_t, h_{t-1}\right] + b_C\right) \tag{27}$$

$$C_t = f_t \times C_{t-1} + i_t \times \overline{C}_t \tag{28}$$

$$o_t = \text{sigmoid}\left(W_o\left[x_t, h_{t-1}\right] + b_o\right) \tag{29}$$

$$h_t = o_t \times \tanh(C_t) \tag{30}$$

where x_t represents the input value of time step t; h_t refers to the output value of time step t; f_t is the forget gate output; W_f and b_f are the weight function and bias value of the forget gate; i_t is the output of the input gate; W_i and b_i are the weight function and bias value of the input gate; $\overline{C}_t$ is the vector of the new candidate values for time step t; W_C and b_C are the weight function and bias value of the tanh layer to calculate $\overline{C}_t$; C_t is the cell state of time step t; o_t is the output of the sigmoid layer; and W_o and b_o are the weight function and bias value of the sigmoid layer.

2.5 The proposed MVMD-LSTM modeling framework

Based on the above-mentioned techniques, a new hybrid method based upon mRMR, MVMD, and LSTM is developed for wind power forecasting in this paper, as depicted in Figs. 5 and 6. The prediction process has been explained in the following.

Step 1. Feature selection The mRMR algorithm is firstly employed to select the most optimal and relevant input feature set from the wind power and its influential factors. Then, the top five features with low redundancy and strong correlation with the wind power are selected to constitute related features.

Step 2. Data decomposition The MVMD algorithm is then employed to comprehensively capture the correlations between the selected factors in the previous step in frequency and time domains and obtain synchronous time–frequency analyses of the factors. Therefore, several sets of smoother sub-series associated with various variables will be acquired.

Step 3. Forecasting model Five sets of the sub-series acquired in the previous step are chosen as the input data.

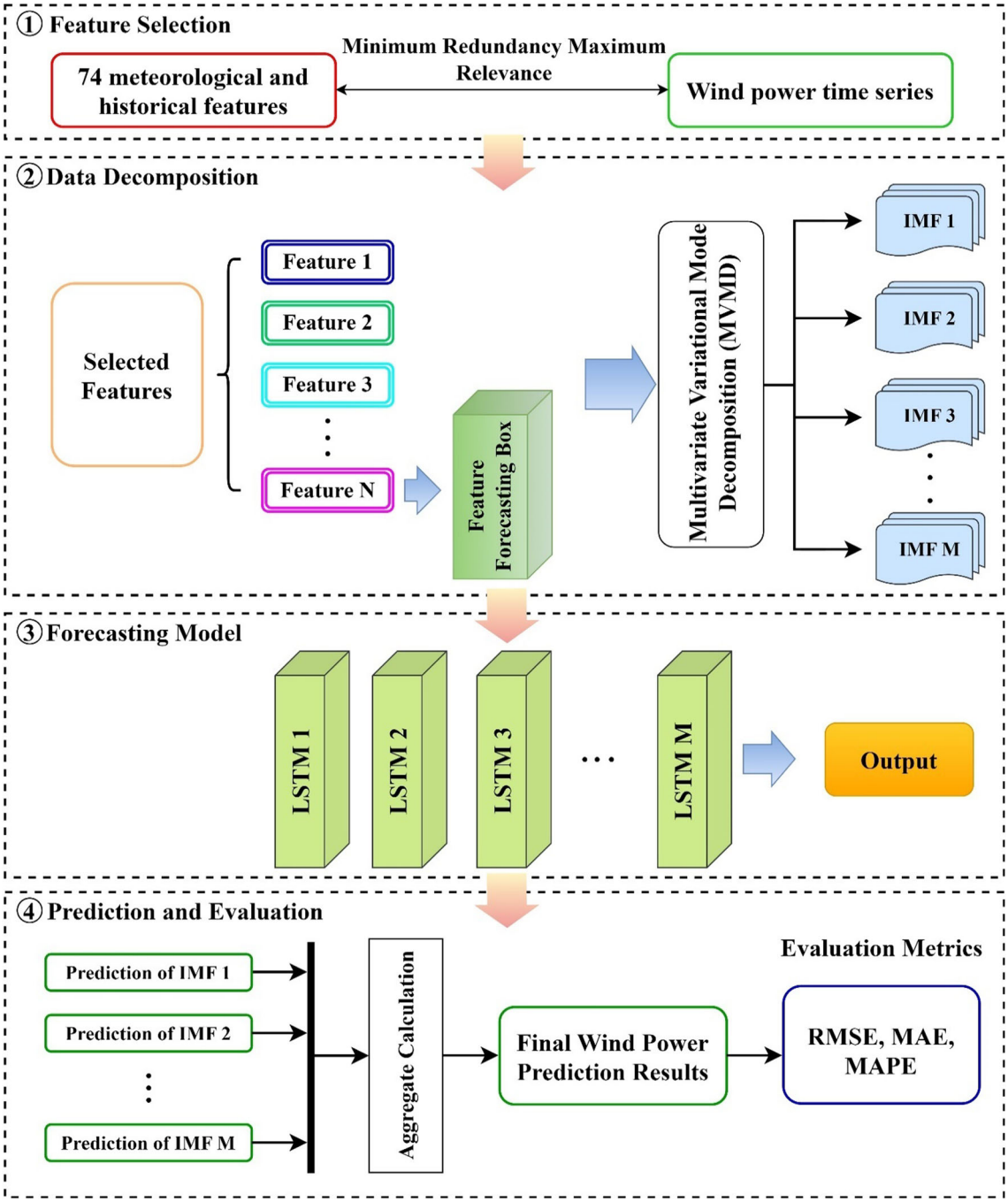

Fig. 5 Framework of the proposed MVMD-LSTM forecasting method

Afterward, these data are used to carry out predictions for the next 24 h based on the LSTM network.

Step 4. Prediction and evaluation Finally, the ultimate prediction results are acquired through summing the corresponding predicted values of each sub-signal, and four evaluation metrics, including RMSE, MAE, MAPE, and R2, are utilized so as to evaluate the forecasting performance.

In Fig. 5, N and M, respectively, indicate the number of selected features and sets of sub-series, and the feature forecasting box illustrated in this figure is as follows.

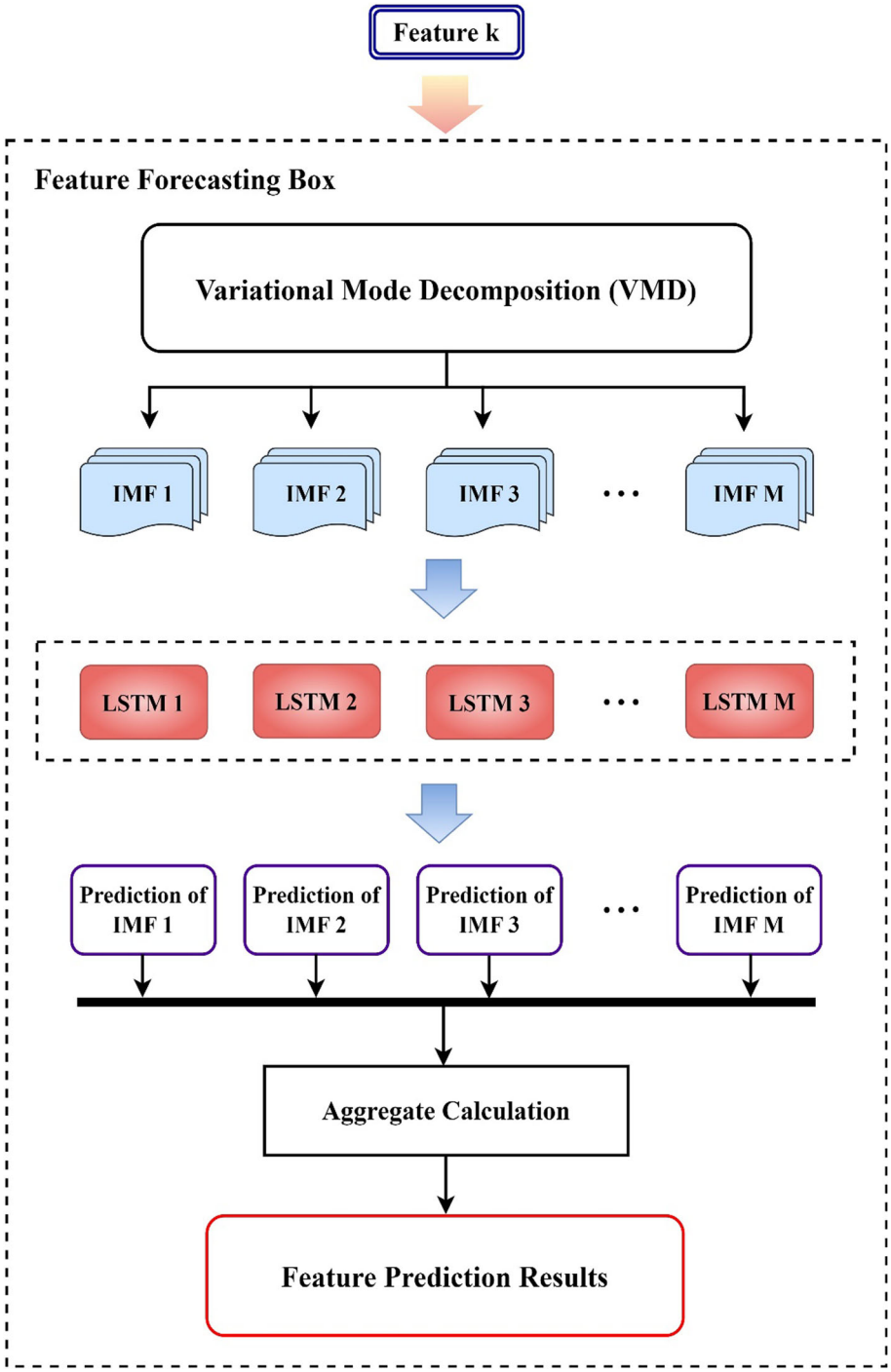

Fig. 6 Internal structure of feature forecasting box in Fig. 5

Table 2 Descriptive statistics of wind power datasets

Dataset	Numbers	Mean (MW)	Max. (MW)	Min. (MW)	Std. (MW)
Alabama					
Samples	4380	20.762	59.918	0	17.622
Washington					
Samples	4380	20.737	55.485	0	21.898

3 Research design

3.1 Data set

Two different real-world practical engineering cases from Alabama (AL) and Washington (WA) with the longitude/latitude of 34.75°N, 87.61°W and 46.006°N, 117.18°W [74] are examined to verify the efficiency of the presented MVMD-LSTM hybrid method. Simulation experiment data sets utilize the 85% of data of ease case as the training samples to construct a model to predict the wind power of the rest. All the related original data are sampled during 6 months of the year 2008 at a 1-h interval. In each series, there are a total number of 4380 data points, with 3722 observations randomly allocated to the training set and the remaining 657 observations designated for the testing set. Descriptive statistics of the wind power dataset are shown in Table 2. It is worth mentioning that in this paper, the resolution of the data is 1 h, and the predicted step length spans 24 h, meaning that the proposed model provides forecasted data for each hour over the next 24-h period. Additionally, given that the data on the forecast day have the closest similarity to those of the previous day, a lag step length of 24 h has been considered. Note that all simulations have been performed on a desktop computer with an Intel (R) Core (TM) i7–4500U CPU @ 2.40 GHz with 6.00 GB of RAM using Python 3.7.

3.2 Evaluation metrics

To assess the prediction accuracy of the developed hybrid method, four estimating indices, root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) have been selected. The expression of each is as follows.

(1) RMSE quantifies the average magnitude of the errors between predicted and actual values, indicating the extent of the difference among them. A low RMSE indicates better model performance.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - \hat{P}_i\right)^2} \tag{31}$$

(2) MAPE is formulated as the following equation. It measures the average percentage error between forecasted and real wind power. Like RMSE, a lower MAPE indicates better model performance.

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - \hat{P}_i}{P_i}\right| \times 100 \tag{32}$$

Table 3 Influencing features and representation variables

Influencing features	Representation variables
Temperature	T
Air pressure	A
Wind direction	D
Wind speed	S
Historical value of temperature	T_t
Historical value of air pressure	A_t
Historical value of wind direction	D_t
Historical value of wind speed	S_t
Historical value of wind power	P_t

(3) MAE is calculated by averaging the absolute differences between real and predicted values.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|P_i - \hat{P}_i\right| \tag{33}$$

where P_i denotes the real value of the wind power; $\hat{P}_i$ refers to the forecasted value of the wind power; $\overline{P}$ represents the average of the real wind power; and N denotes the testing dataset length.

4 Experimental results

4.1 Feature selection analysis

In this paper, first, a feature set is established. Table 3 shows the features and representation variables contained in the set.

The time scale of the wind power in this study is 1 h. As a result, the time t in Table 3 is expressed by 0, 1, …, 24 for all the features (except for P_t whose value at the 24th hour should be predicted, so it includes hours 0–23). The historical values of the features are indicated by the subscript t below them, which indicates the historical value at t hours ago. For example, S_7 represents the wind speed value of the sample 7 h ago. Upon establishing the feature matrix, the candidate feature set J is established for every component. The size of the mRMR value of each feature in the candidate feature set J is determined, and the features are arranged a in descending order according to the magnitude of the mRMR value. The result for each candidate feature set of each feature component is shown in Figs. 7 and 8.

From Figs. 7 and 8, we can see that the top five features with the highest mRMR values have been selected as the related features. These features have low redundancy and high correlation with the wind power. According to these figures, for each case, the mRMR values of wind speed, wind

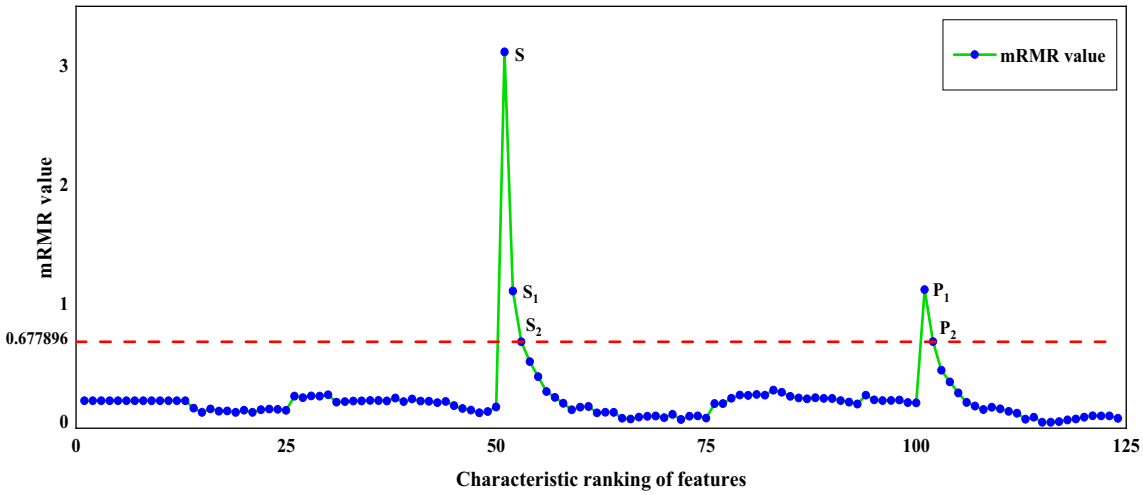

Fig. 7 Feature ranking according to their mRMR values for AL

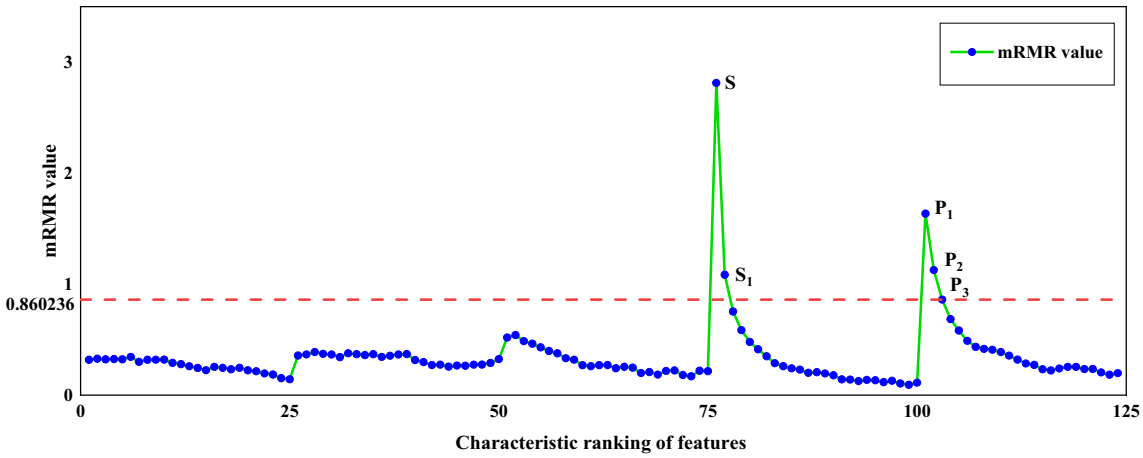

Fig. 8 Feature ranking according to their mRMR values for WA

speed history value, and wind power history value have the highest amount. As for AL, the optimal features are S, S_1, P_1, S_2, and P_2, respectively, and the relevant features for WA are, respectively, S, P_1, P_2, S_1, and P_3.

4.2 Feature prediction

As it is obvious from Figs. 7 and 8, in both cases, the wind speed at the prediction moment (S) has been selected as the most optimal and relevant feature, indicating that it can significantly affect the prediction results. However, we do not have access to wind power at the prediction moment, so we have to estimate its value with high accuracy. In this regard, we have used a VMD-LSTM model to predict S and then used it as a feature so as to forecast wind power, as illustrated in Fig. 6. The hourly historical wind speed data of AL and WA are shown in Fig. 9.

Here, to assess the performance of the VMD-LSTM method in forecasting wind speed, its results have been compared to those of five models, including MLP, SVR, RBF, LSTM, and EMD-LSTM. Figures 10 and 11 show the results for wind speed forecasts of all models for various datasets. The numerical results on the basis of three criteria of MAPE, RMSE, and MAE are represented in Tables 4 and 5.

As seen in Tables 4 and 5, compared to MLP, SVR, and RBF, LSTM tends to show a better performance across all the considered evaluation indices in each case. However, due to the time–frequency decomposition of EMD or VMD, the hybrid models, that is, EMD-LSTM and VMD-LSTM, yield better results than the single models. The mean of the RMSE, MAPE, and MAE of the EMD-LSTM model is 0.95, 11.14, and 0.7, respectively, while the mean of these indices of the VMD-LSTM model is 0.77, 8.9, and 0.61, respectively. Needless to say, in comparison with EMD-LSTM, VMD-LSTM leads to a decrease in MAPE, RMSE, and MAE by 20, 18.9, and 12.9%, respectively. It is clear that VMD-LSTM

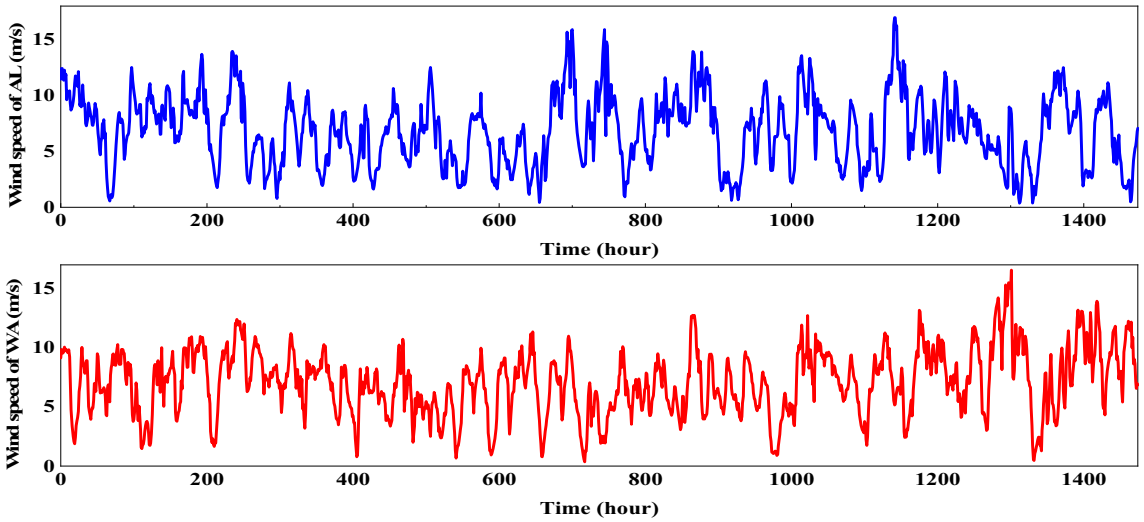

Fig. 9 Original time series of wind speed

outperforms the other models, forecasting wind power with acceptable accuracy.

4.3 Decomposition analysis

Having forecasted wind speed at the prediction moment, we can use it alongside the other selected features in the previous step so as to forecast wind power. At this stage, to more accurately forecast wind power, the multivariate input variables are decomposed all together utilizing the MVMD technique, which is able to take into account the correlations between the variables, thereby reducing prediction difficulty. The original historical wind power data are depicted in Fig. 12, and the decomposition results are illustrated in Fig. 13. It should be noted that since the selected features are wind power, wind speed, and their first, second, and third lags (according to Figs. 7 and 8), their data will be similar (with only a time lag), presenting all of them has been avoided.

Note that the original time series of wind speed is demonstrated in Fig. 9, so they have not been depicted again. In the main signal, the intermittent and complicated wind power nature can be vividly seen. Thus, analyzing wind power time series can come in handy to accurately model its stochastic characteristics.

As demonstrated in Figs. 13 and 14, the decomposition results are seven multivariate imf components ordered from high to low frequency. Modes 1 and 2, which include low-frequency components, properly represent the smooth variations in wind power, while the high-order modes, such as modes 6 and 7, contain higher frequency components. Hence, small fluctuations in wind power can be more accurately detected by high-order decomposed modes. Moreover, it is easy to see that all the multivariate imf components fluctuate up and down at zero, but the imf signal is yet asymmetric. In this paper, to further illustrate the superiority of the MVMD algorithm over common decomposition techniques for wind power forecasting, the proposed MVMD-LSTM model has been compared with EMD-LSTM and VMD-LSTM models for 24-h ahead wind power predictions. The obtained results will be presented and elaborated upon in the next section. The parameters of different models and decomposition techniques are reported in Table 6.

4.4 Statistical analysis

All the models have been applied to the two real-world testing datasets. Then, the RMSE, MAPE, and MAE have been determined for the testing datasets. All the prediction models are executed 10 times independently, generating 10 results of RMSE, MAPE, and MAE for every model. Forecasting accuracy measures of various models for the two testing sets are shown in detail in Tables 7 and 8. The performance is sorted based on the mean of MAPE of the 10 times in AL and WA. In addition, all the prediction results are shown in Figs. 15 and 16.

In this section, to comprehensively evaluate the introduced model's performance, four types of model comparison are proposed. In comparison one, to illustrate the efficacy of the LSTM model in enhancing prediction accuracy, its performance is compared to that of SVR, MLP, and ELM. In comparison two, we compare the performance of different models with and without wind speed prediction, thereby evaluating its significant role in increasing prediction accuracy. In comparison three, LSTM, EMD-LSTM, VMD-LSTM, and MVMD-LSTM have been chosen to demonstrate the advantages of the MVMD algorithm in reducing the forecasting difficulty. In comparison four, the MVMD-LSTM model has

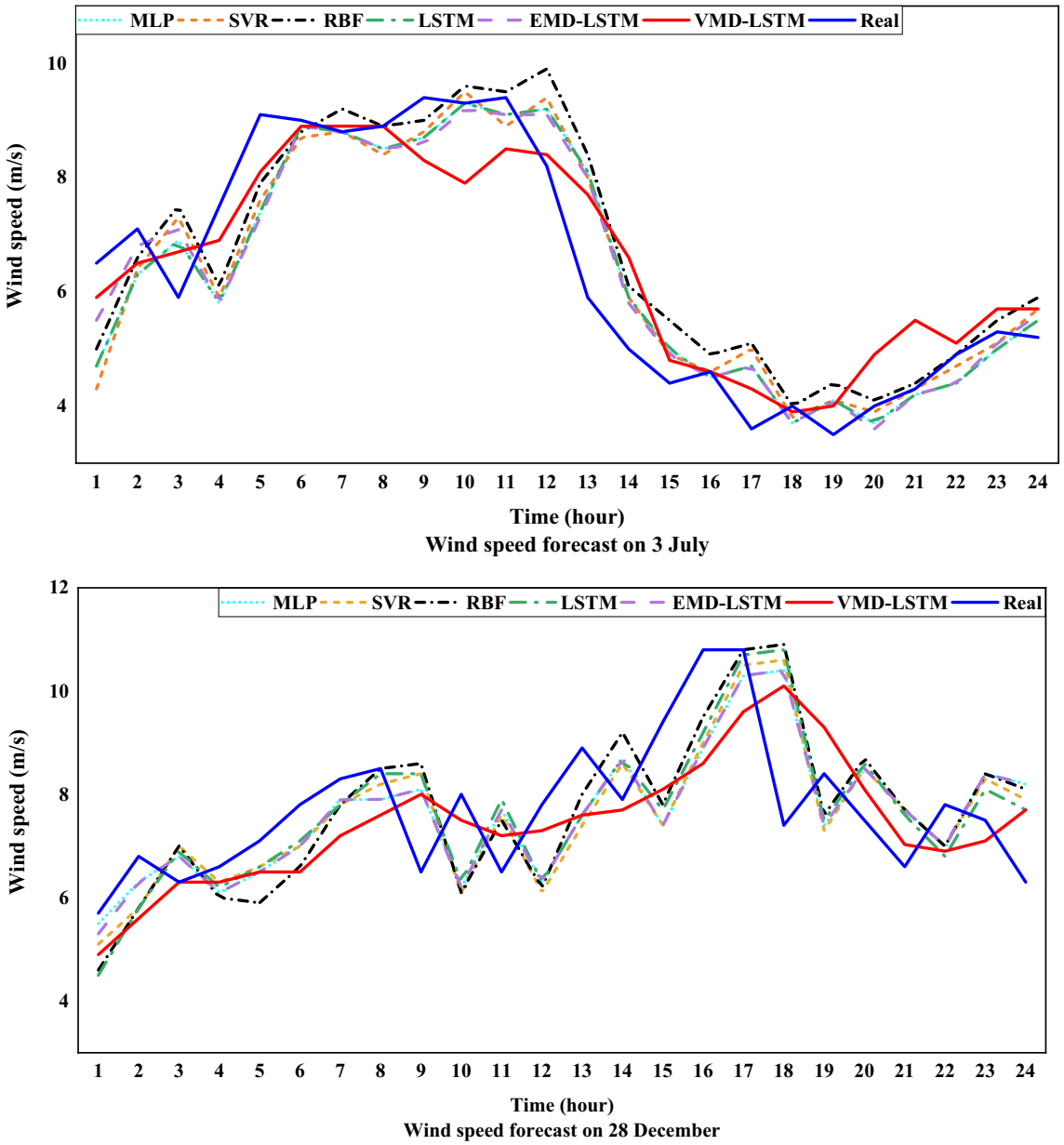

Fig. 10 Wind speed prediction of AL

been comprehensively compared to all the forenamed models.

The comparison results have been summarized as follows.

5 Comparison one

Upon comparing the performance of various single models (SVR, MLP, ELM, and LSTM), it becomes evident that the LSTM model generates superior forecasting results compared to the other models. To be more specific, in comparison with the other single models, LSTM shows the best performance in terms of all metrics, confirming its superiority over the other models and verifying its advantages of enhancing forecasting accuracy. In this regard, LSTM has been chosen as the predictor of the presented method in this paper. Figure 17 compares the performance of the single models.

6 Comparison two

As seen in Figs. 6 and 7, wind speed at the prediction moment was selected by the mRMR feature selection algorithm as the first and best feature with the highest mRMR value. This, in turn, means that this feature has the strongest correlation

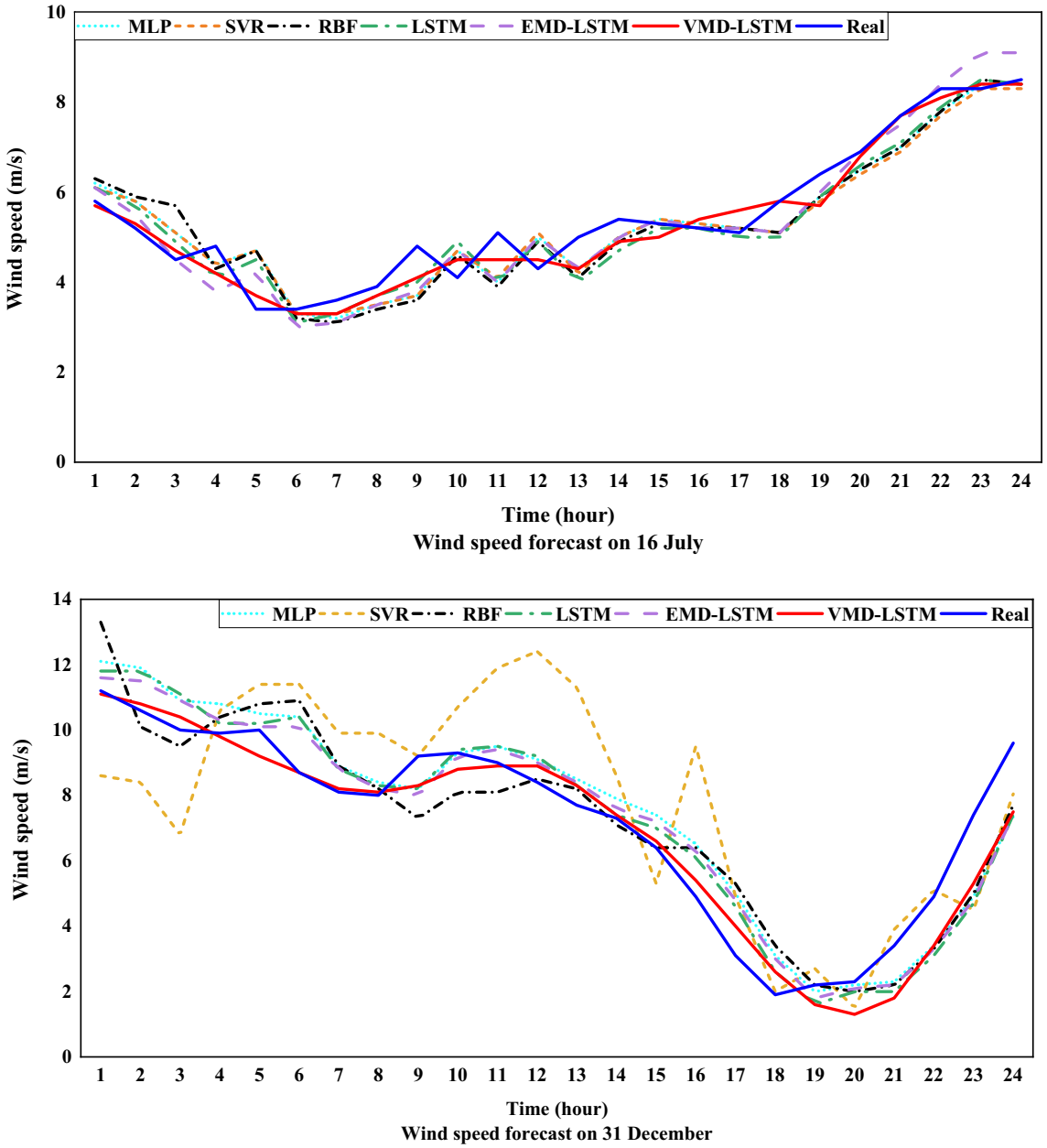

Fig. 11 Wind speed prediction of WA

with wind power, so it does play a crucial role in forecasting results. However, the problem is that we do not have access to wind power at the prediction moment. Some studies have assumed that wind speed at the prediction moment is available, which has nothing to do with reality. As a result, in this paper, we have predicted wind speed at the prediction moment and then used the forecasted values to finally predict wind power. In this comparison, we assess the effect of wind speed prediction at the forecasting time on the results of wind power forecasting. The obtained results with and without predicting wind speed confirm that forecasting it does enhance the wind power prediction accuracy in all the cases. Table 9 represents the difference between various metrics when wind speed is predicted with when it is neglected for the single models. In this table, ΔMAPE, ΔRMSE, and ΔMAE, respectively, refer to the difference between the value of MAPE, RMSE, and MAE when wind speed at the prediction moment is forecasted and when it is not.

According to Table 9, forecasting wind speed and considering it as a feature to predict wind power can reduce MAPE, RMSE, and MAE in all the considered cases, thus improving the accuracy of power forecasting. On the other hand, ignoring this feature adversely affects wind power prediction due to its high correlation with wind power. As a result,

Table 4 Results of evaluation indices for wind speed forecasting obtained by various algorithms for AL

Date	Model	Evaluation indices		
		MAPE	RMSE	MAE
3 July	MLP	10.9	0.94	0.7
	SVR	11.33	0.98	0.73
	RBF	11.61	1.0	0.74
	LSTM	10.8	0.92	0.69
	EMD-LSTM	9.95	0.85	0.64
	VMD-LSTM	**8.68**	**0.67**	**0.56**
28 December	MLP	14.3	1.3	1.1
	SVR	14.63	1.3	1.1
	RBF	15.07	1.3	1.16
	LSTM	14.13	1.28	1.0
	EMD-LSTM	13.86	1.27	1.07
	VMD-LSTM	**12.11**	**1.12**	**0.93**

Bold values indicate the best-performing models

Table 5 Results of evaluation indices for wind speed forecasting obtained by various algorithms for WA

Date	Model	Evaluation indices		
		MAPE	RMSE	MAE
16 July	MLP	9.33	0.61	0.51
	SVR	9.38	0.61	0.51
	RBF	9.73	0.63	0.53
	LSTM	8.95	0.59	0.49
	EMD-LSTM	8.55	0.57	0.47
	VMD-LSTM	**5.62**	**0.38**	**0.30**
31 December	MLP	14.3	1.18	1.0
	SVR	25.18	2.2	1.8
	RBF	13.81	1.2	1
	LSTM	12.56	1.13	0.99
	EMD-LSTM	12.2	1.1	0.88
	VMD-LSTM	**9.0**	**0.89**	**0.65**

Bold values indicate the best-performing models

although it may not be possible to forecast wind speed with 100% accuracy, predicting it can significantly enhance the precision of wind power forecasting.

7 Comparison three

As evident in Tables 7 and 8, in comparison with the LSTM model, the EMD-LSTM and VMD-LSTM models have advantages in terms of forecasting precision. For example, the EMD-LSTM and VMD-LSTM models lead to a decrease in MAPE by 6.82 and 12.43%, respectively, for AL on 3 July. These results verify the efficacy of the mode decomposition techniques in increasing the accuracy of wind power prediction. Furthermore, if we compare the LSTM model with the VMD-LSTM and MVMD-LSTM models, it can be seen that the MVMD-LSTM model achieves higher performance and has the optimal evaluation metrics values in all the considered cases. For instance, the RMSE, MAPE, and MAE errors of MVMD-LSTM are lower than those of VMD-LSTM by 17.63, 11.04, and 8.84%, respectively, for WA on 31 December. Hence, MVMD makes a greater contribution to reducing prediction difficulty and increasing prediction accuracy than VMD, which confirms how important it is to consider the related exogenous variables involved in wind power prediction.

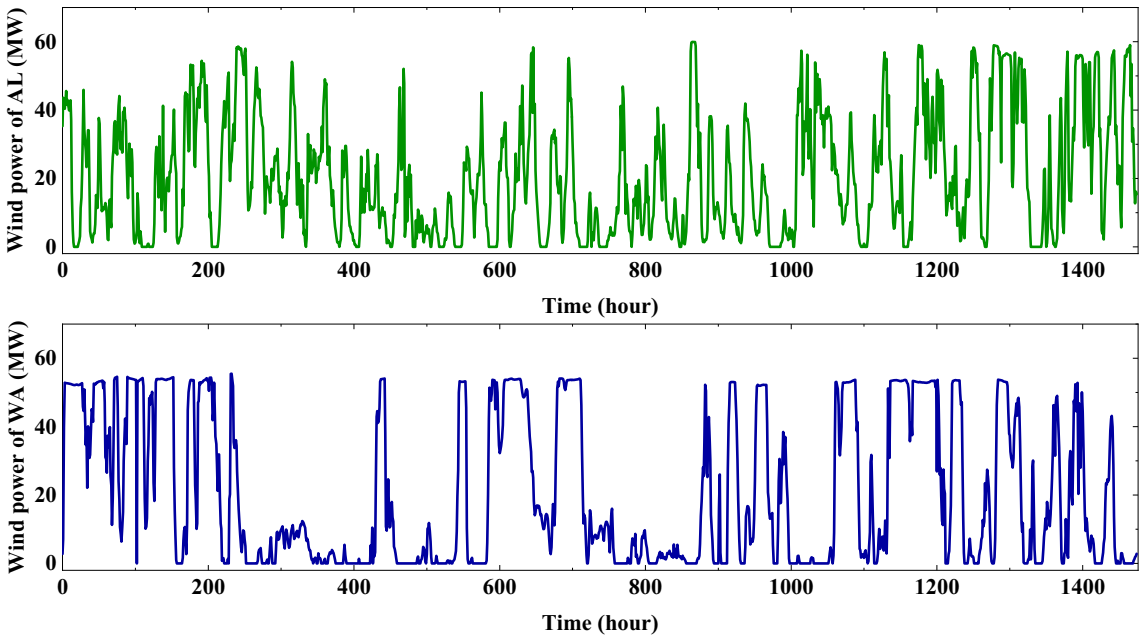

Fig. 12 Original time series of wind power

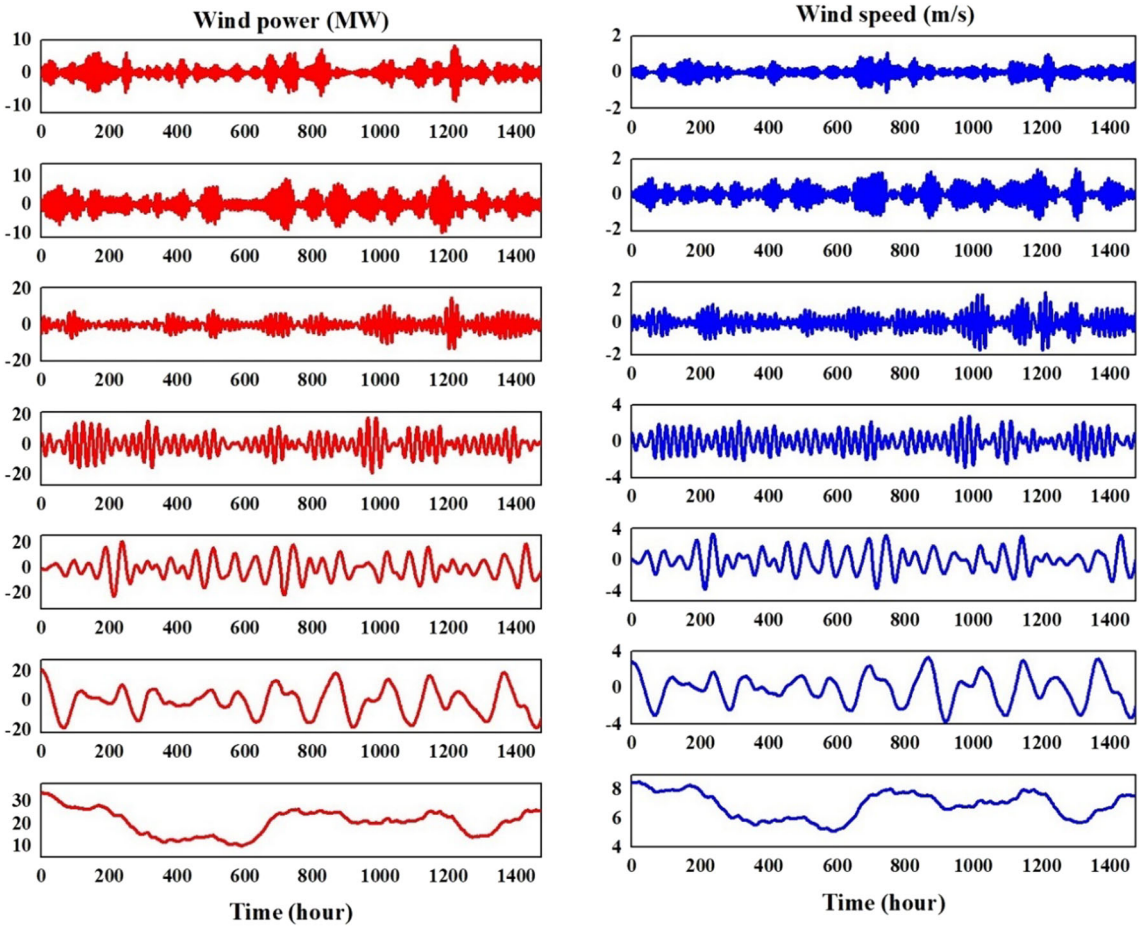

Fig. 13 Decomposition results by MVMD algorithm for AL ($M_{\text{imf}} = 7$)

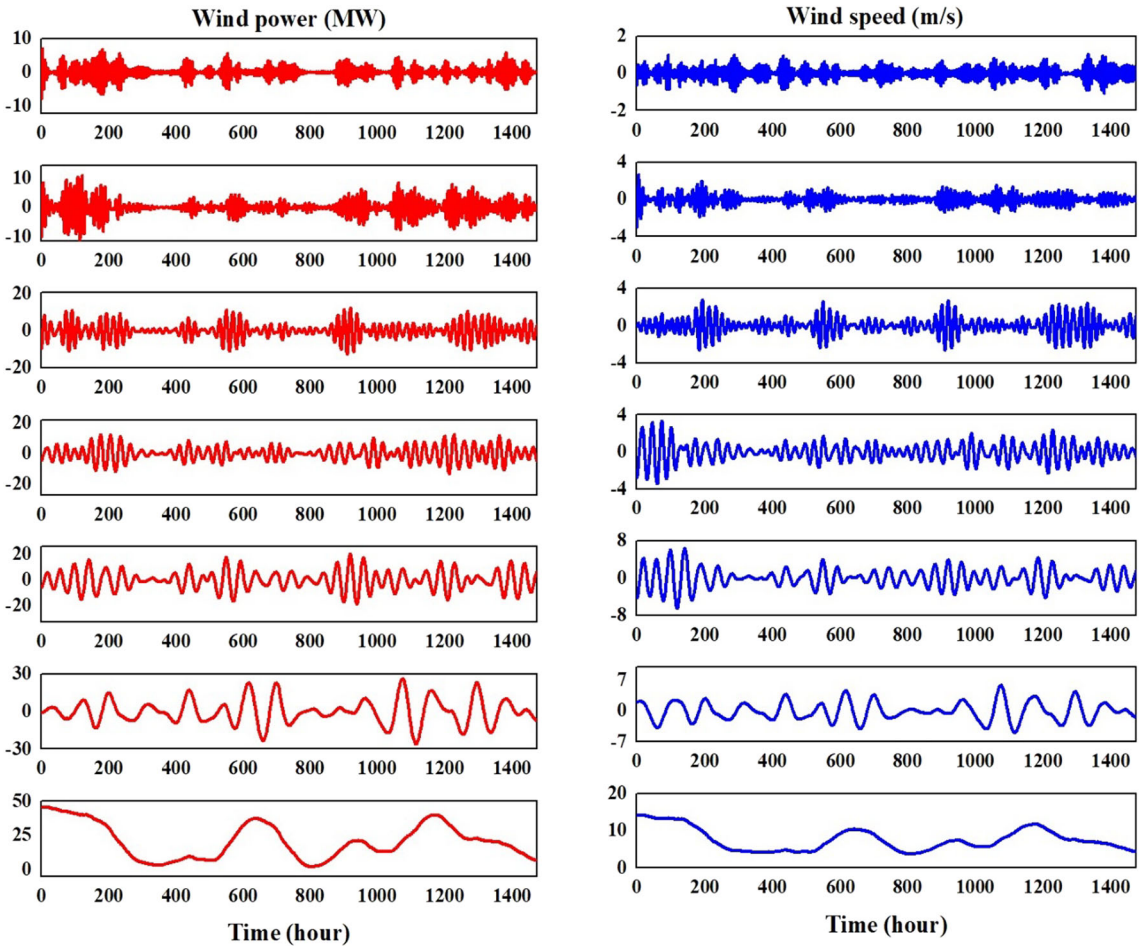

Fig. 14 Decomposition results by MVMD algorithm for WA ($M_{\text{imf}} = 7$)

Table 6 Parameters of models and decomposition techniques

SVR	Penalty factor = 1000; kernel type = polynomial kernel degree = 3; kernel coefficient = 0.1
MLP	Number of hidden layers = 1; number of neurons in hidden layers = 1000; activation function = tanh
ELM	Number of layers = 3; number of neurons in each layer = 24; activation function = tanh; optimizer = Adam
LSTM	number of neurons in each layer = 24; activation function = tanh; optimizer = Adam
VMD	The balancing parameter of the data-fidelity constraint = 2000; the time step of the dual ascent = 0; number of modes = 7; tolerance of convergence criterion = 10^{-7}
EMD	Maximum iteration number = 1000; number of modes = 7
MVMD	The balancing parameter of the data-fidelity constraint = 2000; the time-step of the dual ascent = 0; number of modes = 7; tolerance of convergence criterion = 10^{-7}

Table 7 Wind power forecasting results of the proposed method and the other compared methods for AL

Date	Model	Evaluation indices		
		MAPE	RMSE	MAE
3 July	SVR	40.32	4.999	3.726
	MLP	23.50	4.627	3.457
	ELM	23.42	4.655	3.468
	LSTM	23.01	4.385	3.544
	EMD-LSTM	21.44	4.082	3.169
	VMD-LSTM	20.15	4.164	2.973
	MVMD-LSTM	**19.53**	**4.036**	**2.883**
28 December	SVR	37.14	13.735	10.713
	MLP	36.55	10.118	8.142
	ELM	36.32	10.003	8.091
	LSTM	34.62	10.130	7.712
	EMD-LSTM	34.10	9.538	7.598
	VMD-LSTM	29.69	8.048	6.560
	MVMD-LSTM	**28.57**	**7.747**	**6.312**

Bold values indicate the best-performing models

Table 8 Wind power forecasting results of the proposed method and the other compared methods for WA

Date	Model	Evaluation indices		
		MAPE	RMSE	MAE
16 July	SVR	35.51	2.486	1.876
	MLP	23.32	2.501	1.908
	ELM	21.59	2.270	1.912
	LSTM	21.21	2.252	1.735
	EMD-LSTM	21.11	2.139	1.727
	VMD-LSTM	20.56	2.267	**1.622**
31 December	MVMD-LSTM	**20.49**	**2.106**	1.677
	SVR	25.59	5.583	4.496
	MLP	20.77	5.505	4.720
	ELM	18.30	6.011	4.057
	LSTM	17.33	5.806	3.490
	EMD-LSTM	17.08	6.081	3.786
	VMD-LSTM	16.57	5.489	3.675
	MVMD-LSTM	**14.74**	**4.531**	**3.350**

Bold values indicate the best-performing models

8 Comparison four

According to Tables 7 and 8, the MVMD-LSTM model shows the best performance in day-ahead wind power predictions and has the most favorable values across all evaluation indices, with a minimum MAPE, RMSE, and MAE. Specifically, compared to the second-best model, VMD-LSTM, the MAPE, RMSE, and MAE errors of MVMD-LSTM reduce by 4.19, 7.75, and 4.1% on average, respectively. Thus, the proposed hybrid model manages to achieve considerable prediction accuracy than the other models. Figures 18, 19 and 20 illustrate the scatter plot diagrams, the frequency histograms, and the normal probability plots of the wind power deviations for each model. It can also be seen that the proposed model has the best-fit effect. In addition, the symmetry of the

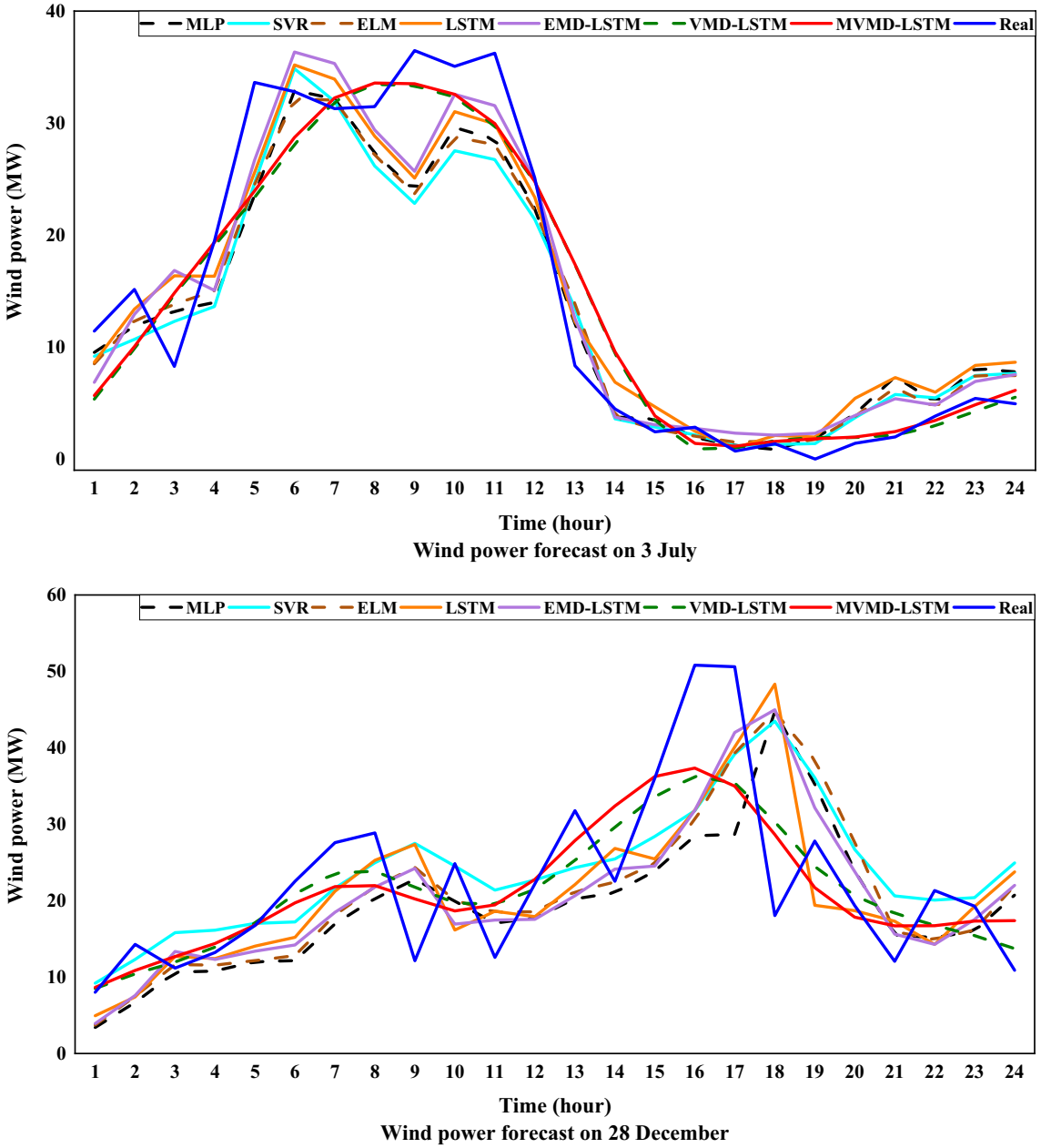

Fig. 15 Wind power prediction of AL

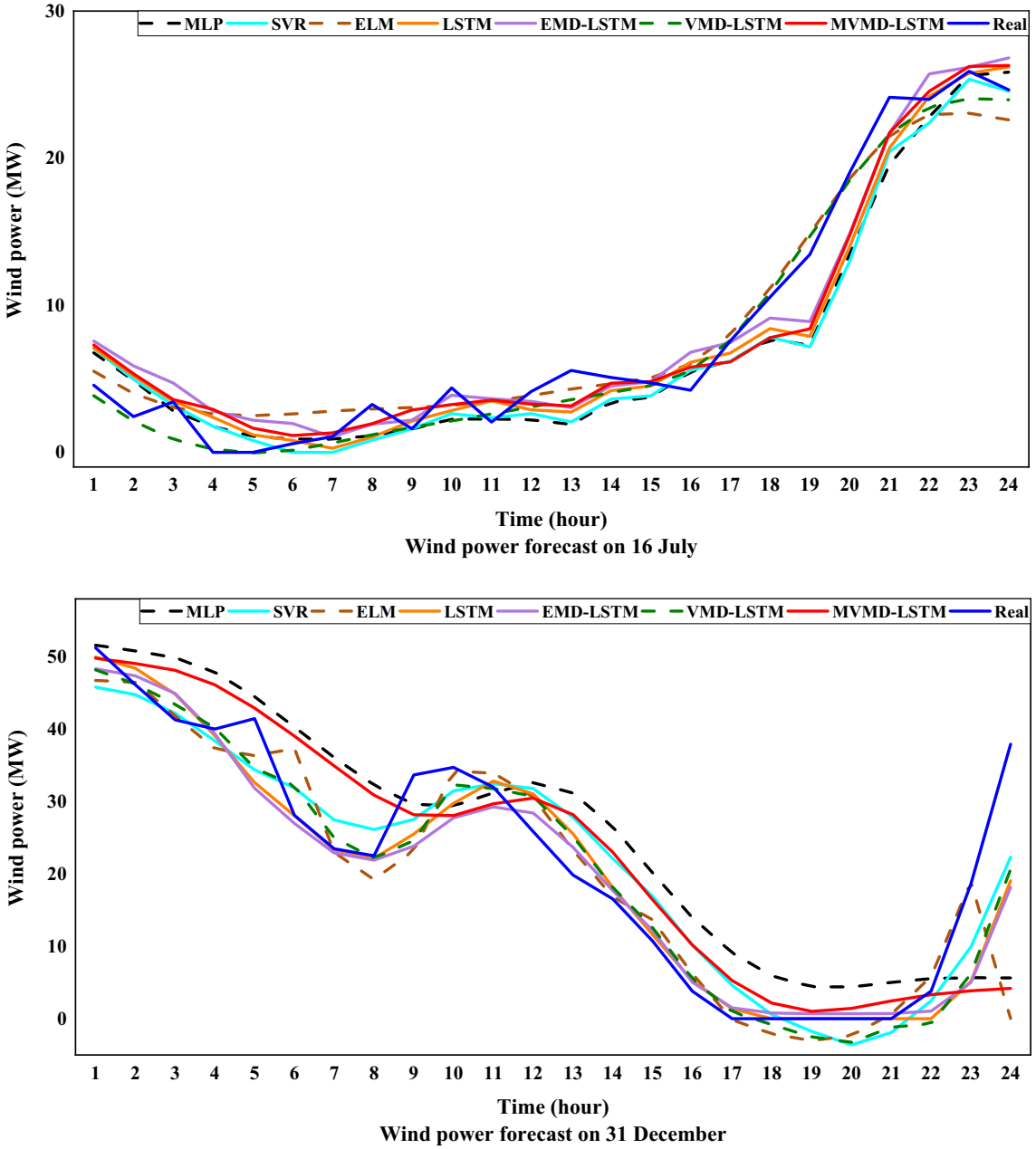

Fig. 16 Wind power prediction of WA

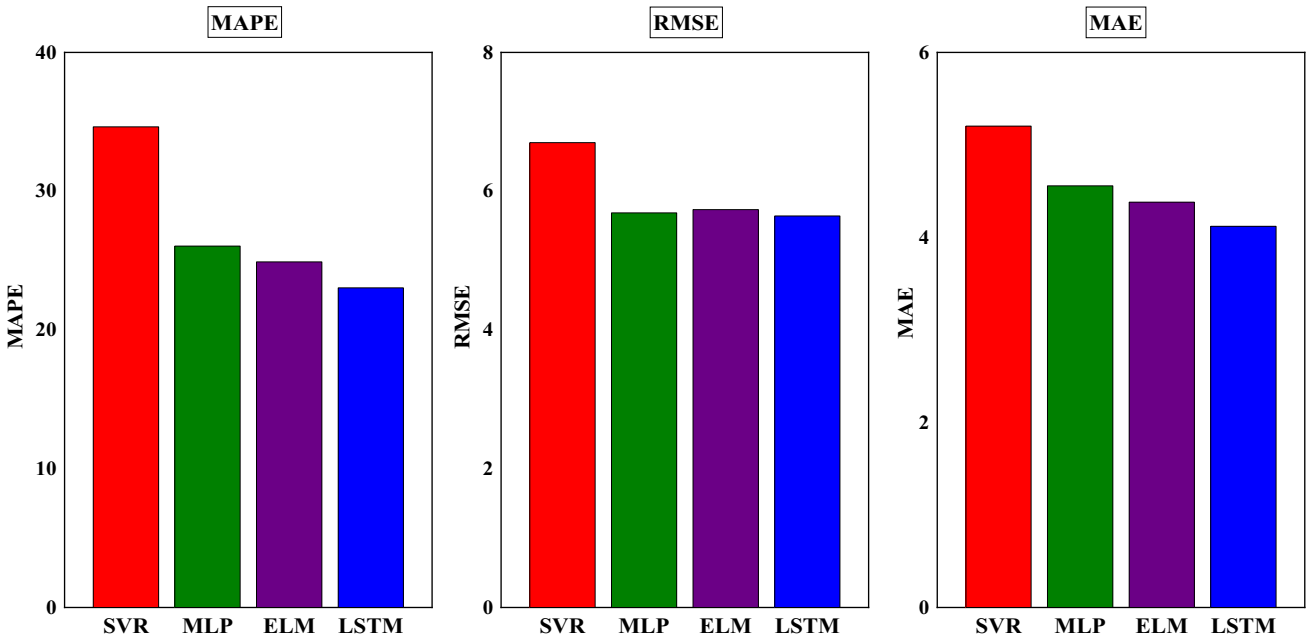

Fig. 17 Comparing the accuracy of wind power forecasting of the single models based on the mean of MAPE, RMSE, and MAE

Table 9 Comparing the prediction accuracy of the single models with and without forecasting wind speed

Model	Metric (%)	Case study			
		AL		WA	
		3 July	28 December	16 July	31 December
SVR	Δ MAPE	70.06	30.93	15.09	3.47
	Δ RMSE	81.19	14.14	4.14	33.58
	Δ MAE	57.56	15.35	12.68	30.58
MLP	Δ MAPE	9.06	17.94	40.82	15.35
	Δ RMSE	34.54	48.64	25.14	18.61
	Δ MAE	9.66	17.95	40.82	12.56
ELP	Δ MAPE	29.24	16.46	16.11	38.68
	Δ RMSE	13.39	29.09	43.37	16.85
	Δ MAE	28.83	16.47	7.26	36.18
LSTM	Δ MAPE	9.30	3.73	13.01	35.60
	Δ RMSE	6.57	2.15	22.30	3.86
	Δ MAE	4.7	3.71	13.02	27.50

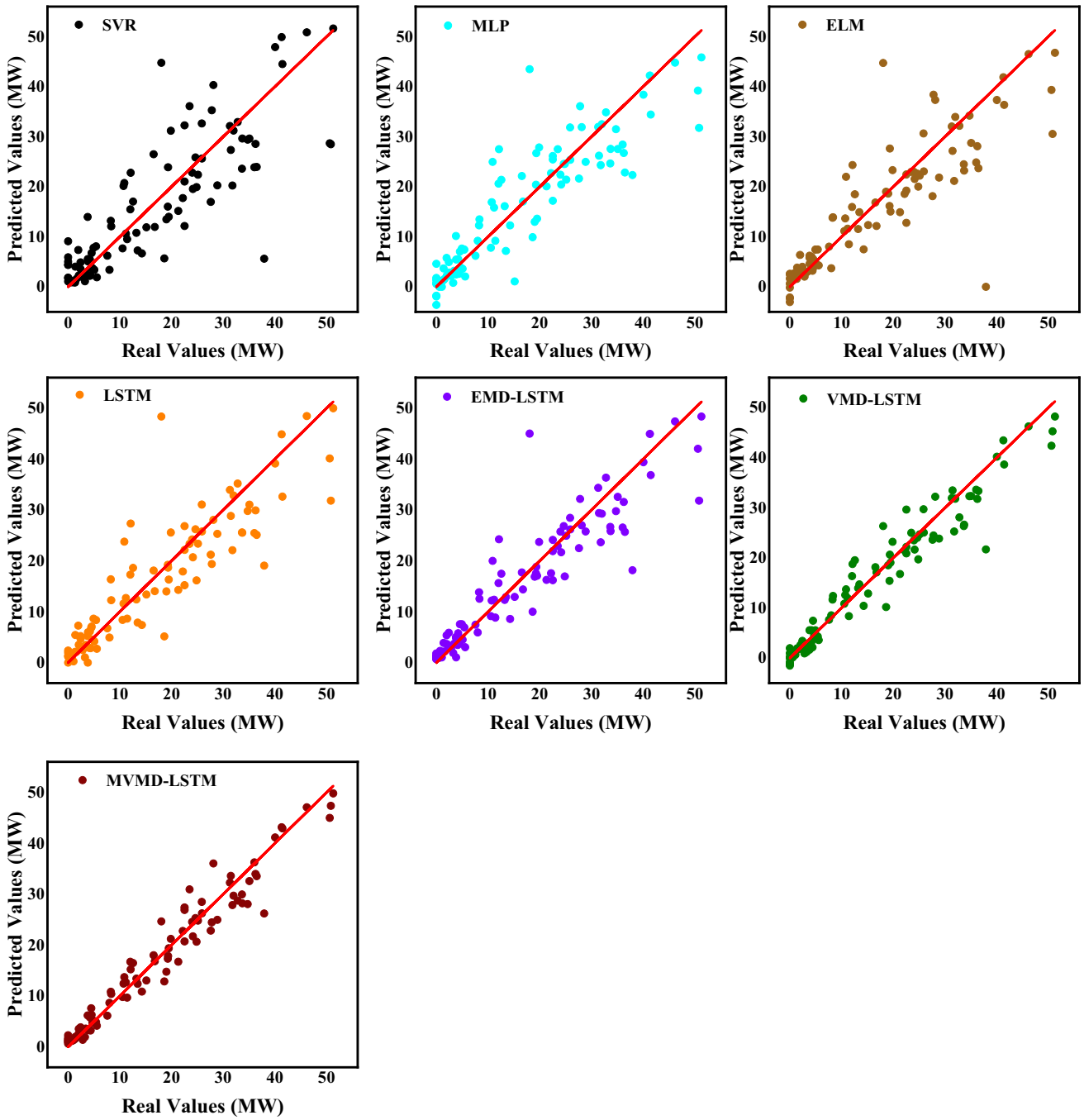

Fig. 18 Scatter plot of prediction and observed values of different prediction models

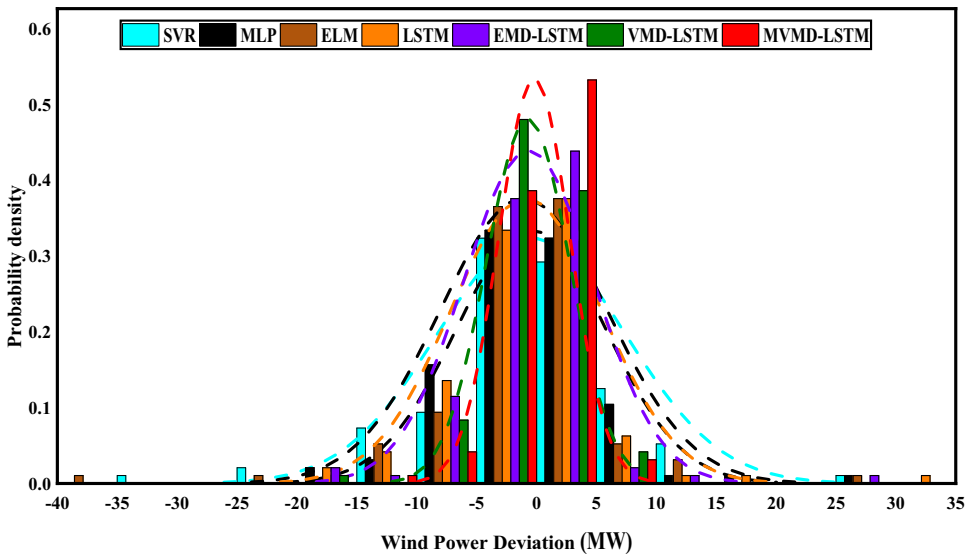

Fig. 19 Frequency histograms of wind power deviations of different prediction models

deviation frequency histogram and the linearity of the normal probability plot collectively indicate that the forecasting model performs well in terms of unbiasedness and normality of deviations. These characteristics validate the robustness of the model and its suitability for reliable wind power forecasting.

9 Conclusion

In this paper, a new two-step forecasting method with outstanding forecasting accuracy is developed for short-term wind power prediction. The proposed procedure starts with selecting the most appropriate features with the strongest correlation with wind power. To do so, the mRMR algorithm is employed; this algorithm selects the best features with high relevance and low redundancy, thereby reducing the input dimension and improving the computation efficiency. Then, those selected features whose values are not available at the prediction moment, such as wind speed, are predicted using a VMD-LSTM model. The results reveal that forecasting such features exerts a unique influence on enhancing the precision of wind power forecasting. In the next step, a multivariate mode decomposition technique, called MVMD, is employed to simultaneously decompose all the selected features. The MVMD technique can decompose the original data into different sub-series so as to mine and capture the major wind power characteristics and the selected features. In this way, it can tackle the frequency mismatches between various series during the decomposition process. The MVMD algorithm outperforms the other two decomposition techniques, that is, EMD and VMD in all cases. Finally, the LSTM neural network is utilized to predict each of the decomposed time series. In the proposed paper, apart from the LSTM neural network, the performance of three neural networks, SVR, MLP, and ELM, is also evaluated. The results illustrate that LSTM leads to the highest accuracy. Furthermore, a comparison between the presented approach and the other five methods on the basis of MAPE, RMSE, and MAE criteria is drawn. Achieving the smallest values of these criteria, the developed approach demonstrates the highest level of performance. Due to the outstanding performance of the presented method, it can make a huge contribution to the grid dispatch plan in real time, reducing the negative effects of wind power fluctuations on the power supply stability. It can be concluded that the developed forecasting model has higher applicability to short-term wind power forecasting. Furthermore, the introduced model is of enormous potential for expansion into other forecasting domains, such as load forecasting as well as photovoltaic generation forecasting.

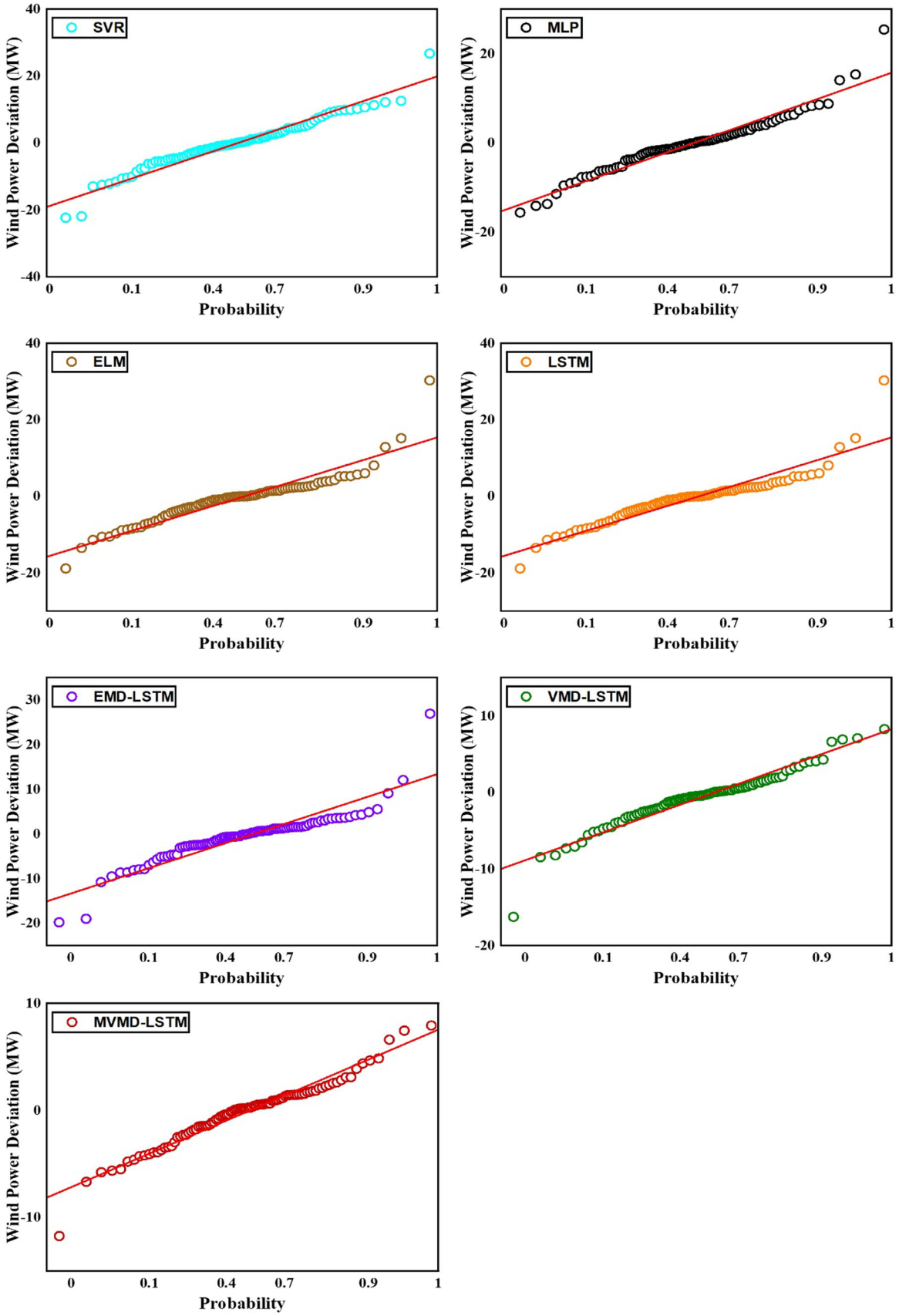

Fig. 20 Normal probability plots of wind power deviations of different prediction models

Author contribution EG involved in conceptualization, methodology, software, formal analysis, visualization, formal analysis, data curation, investigation, validation. AA took part in conceptualization, methodology, software, formal analysis, writing & editing, visualization, data curation, investigation, validation, project administration.

Data availability The wind power data that support the findings of this study can be found here: National Renewable Energy Laboratory. Online. Available: https://www.nrel.gov/

Declaration

Conflict of interest The authors declare no competing interests.

References

1. Ahmed SD, Al-Ismail FS, Shafiullah M, Al-Sulaiman FA, El-Amin IM (2020) Grid integration challenges of wind energy: a review. IEEE Access 8:10857–10878
2. Avar A, Sheikh-El-Eslami MK (2021) Optimal DG placement in power markets from DG Owners' perspective considering the impact of transmission costs. Electric Power Syst Res 196:107218
3. Avar MK (2022) Sheikh-El-Eslami, A new benefit-based transmission cost allocation scheme based on capacity usage differentiation. Electr Power Syst Res 208:107880
4. Tian Z, Li H, Li F (2021) A combination forecasting model of wind speed based on decomposition. Energy Rep 7:1217–1233
5. Qiao B, Liu J, Wu P, Teng Y (2022) Wind power forecasting based on variational mode decomposition and high-order fuzzy cognitive maps. Appl Soft Comput 129:109586
6. Zhu J, Su L, Li Y (2022) Wind power forecasting based on new hybrid model with TCN residual modification. Energy and AI 10:100199
7. Dou C, Zheng Y, Yue D, Zhang Z, Ma K (2018) Hybrid model for renewable energy and loads prediction based on data mining and variational mode decomposition. IET Gener Transm Distrib 12(11):2642–2649
8. Cassola F, Burlando M (2012) Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. Appl Energy 99:154–166
9. De Giorgi MG, Ficarella A, Tarantino M (2011) Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy 36(7):3968–3978
10. Li J, Zhang S, Yang Z (2022) A wind power forecasting method based on optimized decomposition prediction and error correction. Electric Power Syst Res 208:107886
11. Do D-PN, Lee Y, Choi J (2016) Hourly average wind speed simulation and forecast based on ARMA model in Jeju Island, Korea. J Electr Eng Technol 11(6):1548–1555
12. Sim S-K, Maass P, Lind PG (2018) Wind speed modeling by nested ARIMA processes. Energies 12(1):69
13. Lydia M, Kumar SS, Selvakumar AI, Kumar GEP (2016) Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers Manag 112:115–124
14. Erdem E, Shi J (2011) ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 88(4):1405–1414
15. Shukur OB, Lee MH (2015) Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. Renew Energy 76:637–647
16. Kavasseri RG, Seetharaman K (2009) Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 34(5):1388–1393
17. Yuan X, Tan Q, Lei X, Yuan Y, Wu X (2017) Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 129:122–137
18. Sun G et al (2018) Short-term wind power forecasts by a synthetical similar time series data mining method. Renew Energy 115:575–584
19. Yueheng W (2021) Short-term prediction of wind power based on Kalman filter tracking fusion. In: 2021 4th International conference on energy, electrical and power engineering (CEEPE), pp 507–511: IEEE
20. Chen N, Qian Z, Nabney IT, Meng X (2013) Wind power forecasts using Gaussian processes and numerical weather prediction. IEEE Trans Power Syst 29(2):656–665
21. Wang Y, Hu Q, Meng D, Zhu P (2017) Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. Appl Energy 208:1097–1112
22. Liu L et al (2023) Ultra-short-term wind power forecasting based on deep Bayesian model with uncertainty. Renew Energy 205:598–607
23. Hu H, Wang L, Tao R (2021) Wind speed forecasting based on variational mode decomposition and improved echo state network. Renew Energy 164:729–751
24. Fu W, Fang P, Wang K, Li Z, Xiong D, Zhang K (2021) Multi-step ahead short-term wind speed forecasting approach coupling variational mode decomposition, improved beetle antennae search algorithm-based synchronous optimization and Volterra series model. Renew Energy 179:1122–1139
25. Zeng J, Qiao W (2012) Short-term wind power prediction using a wavelet support vector machine. IEEE Trans Sustain Energy 3(2):255–264
26. Wang J, Fang K, Pang W, Sun J (2017) Wind power interval prediction based on improved PSO and BP neural network. J Electr Eng Technol 12(3):989–995
27. Khazaei S, Ehsan M, Soleymani S, Mohammadnezhad-Shourkaei H (2022) A high-accuracy hybrid method for short-term wind power forecasting. Energy 238:122020
28. Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput 31(7):1235–1270
29. Guo Z, Zhao W, Lu H, Wang J (2012) Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 37(1):241–249
30. Wang K, Fu W, Chen T, Zhang B, Xiong D, Fang P (2020) A compound framework for wind speed forecasting based on comprehensive feature selection, quantile regression incorporated into convolutional simplified long short-term memory network and residual error correction. Energy Convers Manage 222:113234
31. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
32. Martínez-Martínez JM, Escandell-Montero P, Soria-Olivas E, Martín-Guerrero JD, Magdalena-Benedito R, Gómez-Sanchis J (2011) Regularized extreme learning machine for regression problems. Neurocomputing 74(17):3716–3721

33. Zhao X, Wei H, Li C, Zhang K (2020) A hybrid nonlinear forecasting strategy for short-term wind speed. Energies 13(7):1596
34. Wang J, Wang S, Yang W (2019) A novel non-linear combination system for short-term wind speed forecast. Renew Energy 143:1172–1192
35. Zhou Q, Lv Q, Zhang G (2021) A combined forecasting system based on modified multi-objective optimization for short-term wind speed and wind power forecasting. Appl Sci 11(20):9383
36. Duan J, Wang P, Ma W, Fang S, Hou Z (2022) A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting. Int J Electr Power Energy Syst 134:107452
37. Li Z, Li C (2018) Non-Gaussian non-stationary wind pressure forecasting based on the improved empirical wavelet transform. J Wind Eng Ind Aerodyn 179:541–557
38. Qian Z, Pei Y, Zareipour H, Chen N (2019) A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Appl Energy 235:939–953
39. Bhaskar K, Singh SN (2012) AWNN-assisted wind power forecasting using feed-forward neural network. IEEE Trans Sustain Energy 3(2):306–315
40. Liu D, Niu D, Wang H, Fan L (2014) Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 62:592–597
41. Wang LJ, Dong L, Hu GF, Gao S, Liao XZ (2010) Combined prediction of wind power generation in multi-dimension embedding phase space. Cntrl Decision 25(4)
42. Han L, Romero CE, Yao Z (2015) Wind power forecasting based on principle component phase space reconstruction. Renew Energy 81:737–744
43. Jiang Y, Liu S, Zhao N, Xin J, Wu B (2020) Short-term wind speed prediction using time varying filter-based empirical mode decomposition and group method of data handling-based hybrid model. Energy Convers Manage 220:113076
44. Yang H-F, Chen Y-PP (2019) Representation learning with extreme learning machines and empirical mode decomposition for wind speed forecasting methods. Artif Intell 277:103176
45. Bokde N, Feijóo A, Al-Ansari N, Tao S, Yaseen ZM (2020) The hybridization of ensemble empirical mode decomposition with forecasting models: Application of short-term wind speed and power modeling. Energies 13(7):1666
46. Yu M (2020) Short-term wind speed forecasting based on random forest model combining ensemble empirical mode decomposition and improved harmony search algorithm. Int J Green Energy 17(5):332–348
47. Tian Z, Li S, Wang Y (2020) A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed. Wind Energy 23(2):177–206
48. Wu C, Wang J, Chen X, Du P, Yang W (2020) A novel hybrid system based on multi-objective optimization for wind speed forecasting. Renew Energy 146:149–165
49. Zhang G et al (2019) Wind power prediction based on variational mode decomposition multi-frequency combinations. J Mod Power Syst Clean Energy 7(2):281–288
50. Abdoos AA (2016) A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. Neurocomputing 203:111–120
51. He F, Zhou J, Feng Z-K, Liu G, Yang Y (2019) A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Appl Energy 237:103–116
52. Han L, Zhang R, Wang X, Bao A, Jing H (2019) Multi-step wind power forecast based on VMD-LSTM. IET Renew Power Gener 13(10):1690–1700
53. Duan J et al (2021) Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and Correntropy Long Short-term memory neural network. Energy 214:118980
54. Dai Y, Zhang M, Xin X, Chen X, Li Y (2024) Short-term wind speed forecasts through hybrid model based on improved variational mode decomposition. Available at SSRN 4312832
55. Han L, Jing H, Zhang R, Gao Z (2019) Wind power forecast based on improved Long Short Term Memory network. Energy 189:116300
56. Prasad R, Ali M, Kwan P, Khan H (2019) Designing a multistage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. Appl Energy 236:778–792
57. Huang Y, Hasan N, Deng C, Bao Y (2022) Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting. Energy 239:122245
58. Jamei M, Ali M, Karbasi M, Xiang Y, Ahmadianfar I, Yaseen ZM (2022) Designing a Multi-Stage Expert System for daily ocean wave energy forecasting: A multivariate data decomposition-based approach. Appl Energy 326:119925
59. Zheng Z et al (2023) Design data decomposition-based reference evapotranspiration forecasting model: a soft feature filter based deep learning driven approach. Eng Appl Artif Intell 121:105984
60. Jamei M et al (2023) A high dimensional features-based cascaded forward neural network coupled with MVMD and Boruta-GBDT for multi-step ahead forecasting of surface soil moisture. Eng Appl Artif Intell 120:105895
61. Zhang K, Yang X, Wang T, Thé J, Tan Z, Yu H (2023) Multi-step carbon price forecasting using a hybrid model based on multivariate decomposition strategy and deep learning algorithms. J Clean Prod 405:136959
62. Xue Y et al (2016) Adaptive ultra-short-term wind power prediction based on risk assessment. CSEE J Power Energy Syst 2(3):59–64
63. Wang J, Han X, Jiang J, Li W, Ma Y (2017) Short-term wind power probabilistic forecasting considering spatial correlation. In: 2017 IEEE conference on energy internet and energy system integration (EI2), pp 1–6: IEEE
64. Buhan S, Çadırcı I (2015) Multistage wind-electric power forecast by using a combination of advanced statistical methods. IEEE Trans Industr Inf 11(5):1231–1242
65. Sharma A, Paliwal KK, Imoto S, Miyano S (2014) A feature selection method using improved regularized linear discriminant analysis. Mach Vis Appl 25:775–786
66. Zhao H-X, Magoulès F (2012) Feature selection for predicting building energy consumption based on statistical learning method. J Algor Comput Technol 6(1):59–77
67. Wang Y, Zou R, Liu F, Zhang L, Liu Q (2021) A review of wind speed and wind power forecasting with deep neural networks. Appl Energy 304:117766
68. Dragomiretskiy K, Zosso D (2013) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–544
69. Lin Z, Sun M, Wu X (2022) Detecting and diagnosing process nonlinearity-induced unit-wide oscillations based on an optimized multivariate variational mode decomposition method. IEEE Access 10:36106–36122
70. Memarzadeh G, Keynia F (2021) Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. Electr Power Syst Res 192:106995

71. Avar A, Ghanbari E (2024) Optimal integration and planning of PV and wind renewable energy sources into distribution networks using the hybrid model of analytical techniques and metaheuristic algorithms: A deep learning-based approach. Comput Electr Eng 117:109280
72. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
73. Abdel-Nasser M, Mahmoud K (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput Appl 31:2727–2740
74. National Renewable Energy Laboratory. Online. Available: https://www.nrel.gov/