Applied Energy 230 (2018) 429–443

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy

Jujie Wang[a,*], Yaning Li[b]

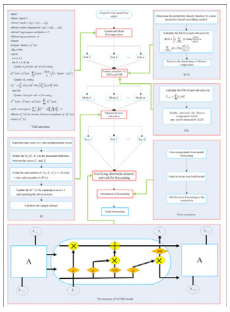
[a] Climate and Weather Disasters Collaborative Innovation Center, School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

[b] College of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

HIGHLIGHTS

- Develop an optimal feature extraction algorithm to capture the optimal features of wind speed fluctuations.
- Introduce an error correction strategy to improve the prediction precision of wind speed.
- A innovate hybrid model is successfully proposed for multi-step ahead wind speed prediction.
- Design three experiments from the real wind farms to validate the availability and reliability of the developed model.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Multi-step ahead wind speed prediction
Optimal feature extraction
Long short term memory network
Error correction strategy
Hybrid model

ABSTRACT

Forecasting wind speed accurately is a key task in the planning and operation of wind energy generation in power systems, and its importance increases with the high integration of wind power into the electricity market. This research proposes an innovative hybrid model based on optimal feature extraction, deep learning algorithm and error correction strategy for multi-step wind speed prediction. The optimal feature extraction including variational mode decomposition, Kullback-Leibler divergence, energy measure and sample entropy is utilized to catch the optimal features of wind speed fluctuations for balancing the calculation efficiency and prediction accuracy. The deep learning algorithm based on long short term memory network, is utilized to detect the long-term and short-term memory characteristics and build the suitable prediction model for each feature sub-signal. The error correction strategy based on a Generalized auto-regressive conditionally heteroscedastic model is developed to modify the above prediction errors when its inherent correlation and heteroscedasticity cannot be ignored. Three real forecasting cases are applied to test the performance and effectiveness of the developed model. The simulation results indicate that the developed model consistently has the smallest statistical errors, and outperforms other benchmark methods. It can be concluded that the developed approach is conductive to strengthening the prediction precision of wind speed.

1. Introduction

Wind power being one of the most promising renewable energy sources, has a rapid development throughout the world recently. The total installed capacity of wind power has been doubled in the past 3 years, and it is estimated that in 2020, approximately 12% of total

* Corresponding author.
E-mail address: jujiewang@126.com (J. Wang).

https://doi.org/10.1016/j.apenergy.2018.08.114
Received 5 April 2018; Received in revised form 11 August 2018; Accepted 20 August 2018
Available online 29 August 2018

world electricity demands will be supplied from wind power [1]. With the high integration of wind power into electric power, the power system is becoming more unreliable because of the intermittent and stochastic natures of wind speed fluctuations [2]. To guarantee a reliable system operation, the power system operator has to schedule sufficient spinning reserve for wind power, which will increase the operating cost of wind power and limit the large-scale exploitation and utilization of wind energy [3]. Thus, forecasting wind speed accurately is indispensable for reducing the operating cost of wind power and enhancing the efficiency of wind power utilization [4].

In the past few decades, lots of methods have been developed for wind speed forecasting, which are usually classified into two categories including physical methods and statistical methods [5]. Physical methods employ the meteorological parameters and physical laws to establish the mathematical models for wind speed forecasting, which usually require substantial computational time and are not good at short-term prediction [6]. Statistical methods intend to use the historical samples to simulate the wind speed fluctuations, which have advantages for short-term prediction. These statistical methods are usually divided into three groups including time series models, machine learning models and hybrid models [7]. These time series models include auto-regressive (AR) model, moving average model (MA) and auto-regressive moving average model (ARMA). Erdem and Shi [8] adopted four ARMA models to predict the vector of wind speed and direction, and the simulate results validated the effectiveness of these models. Lydia et al. [9] used several AR models to enhance the prediction performance of wind velocity. Cadenas and Rivera [10] predict the wind speed by an improved ARMA model, and the results showed that the prediction precision of the proposed model was higher than the persistence model. Although these time series models have satisfactory forecasting performance when wind speed signal shows linearity and stationarity, but they suffer from the disadvantages of nonlinear fitting capability weakness because of their linear assumption among time series. To capture these nonlinear characteristics of wind speed change, lots of machine learning models have been proposed to perform wind speed prediction. For instance, Ren et al. [11] introduced a particle swarm optimization (PSO) to improve the prediction performance of a back propagation (BP) neural network, and the simulate results demonstrated the effectiveness of the proposed model. Zhang et al. [12] presented a radial basis function (RBF) neural network for improving the prediction performance of wind speed. Santamaría-Bonfil et al. [13] employed a support vector regression (SVR) to conduct the wind speed prediction, and the simulate results indicated that the SVR model outperformed these benchmark models. In addition, numerous hybrid models have been developed for wind speed forecasting. For instance, Salcedo-Sanz et al. [14] presented a hybrid model based on fifth generation mesoscale model and artificial neural networks (ANNs) for wind speed forecasting. Song et al. [15] developed a novel hybrid model based on advanced optimization algorithm to improve the forecasting performance of wind speed. Wang et al. [16] proposed a hybrid system based on multi-objective whale optimization algorithm for wind speed prediction. Zhao et al. [17] developed a novel hybrid model based on a weather research and forecasting (WRF) model and an optimized association approach for multi-step ahead wind speed and power prediction. Salcedo-Sanz et al. [18] exploited the input data diversity and developed a novel hybrid model based on physical models and ANNs to improve the forecasting performance of wind speed.

Recently, numerous feature selection methods have been introduced to enhance the forecasting ability of the mainstream prediction models. These methods are usually classified into two categories including wrapper methods and filter methods [19]. Wrapper methods may require large amount of computing time and are usually combined with fast machine learning prediction models, whereas filter methods are performed based on the data and are usually faster than wrapper methods [20]. Thus, the filter methods mainly including wavelet decomposition (WD) and empirical mode decomposition (EMD), have been widely applied in wind speed prediction issues in recent years. For instance, Kiplangat et al. [21] utilized the WD to select the features from original wind speed signal and build the ARMA model for prediction. The results showed the advantages of feature selection method based on WD. Zhang et al. [22] developed two feature selection-based hybrid models which combined EMD with machine learning models (ANN and SVR) for wind speed forecasting, and the simulate results validated the effectiveness of the proposed models. In general, the WD technique has advantages for time-frequency analysis [23], while the EMD has better self-adaptability in handling the chaotic nature and inherent complexity of original signal [22]. Nevertheless, there are some drawbacks: (a) the performance of the WD relies on the choice of wavelet basis and decomposition levels highly; (b) the EMD lacks the clear physical meaning and strict mathematical theory; (c) these common WD (or EMD)-based hybrid models may be unreasonable and cause a big disturbance on the final forecasting because not all sub-signals obtained by WD (or EMD) are beneficial in wind speed prediction. To overcome these disadvantages, it is necessary to develop an effective feature selection algorithm for extraction meaningful features and removing the irrelevant features from original wind speed signal.

The prediction models are the core part of these feature selection-based hybrid models. Different from the shallow learning models, the deep learning models, such as deep belief network (DBN), convolutional neural network (CNN) and recurrent neural network (RNN), can capture the deep inherent features from original data and have been developed rapidly in recent years [24]. Kuremoto et al. [25] designed a DBN with restricted Boltzmann machines for time series prediction and the simulate results validated the effectiveness of the proposed model. Wang et al. [26] presented a CNN model for probabilistic wind power prediction and concluded that the proposed model was superior to the benchmark models. However, these deep learning models are not widely used for wind speed prediction. Taking into account the long-term and short-term dependency of wind speed change, a special kind of RNN called as long short term memory network (LSTMN), is utilized to detect the long-term and short-term memory natures of wind speed change in this study.

Additionally, recent studies show that the error correction strategy (ECS) is also one of the most effective ways to improve the prediction accuracy of wind speed [27]. In Ref. [28], a Markov model was used to correct the prediction errors of the SVR and the results verified the effectiveness of the ECS. Shi et al. [29] employed ANN and SVR to correct the forecasting errors of the ARMA and the simulate results validated the contribution of the ECS. However, most of these filter-based methods neglect the errors of the prediction models because of the hypothesis of white noise. It is obviously not true.

In this study, a novel filter-based hybrid model is developed for multi-step ahead wind speed prediction, which combines deep learning algorithm (DLA) with optimal feature extraction (OFE) and ECS. The OFE includes variational mode decomposition (VMD), Kullback-Leibler divergence (KLD), energy measure (EM) and sample entropy (SE). The proposed model is composed of five steps as follows: (a) the VMD is utilized to resolve a non-stationary wind speed signal into several more stationary sub-signals; (b) two feature selection algorithms including KLD and EM are applied to capture meaningful features from original wind speed signal and remove the disturbance of illusive components introduced by filter algorithm itself; (c) the SE is adopted to recombine the features obtained from two feature selection algorithms for balancing the calculation efficiency and prediction accuracy; (d) the LSTMN is employed to establish the prediction model for each feature sub-signal; (e) the hybrid of LSTMN and Generalized auto-regressive conditionally heteroscedastic (GARCH) is adopted to correct the above prediction errors when its inherent correlation and heteroscedasticity cannot be neglected. Three real forecasting cases are applied to verify the performance and effectiveness of the proposed model. The simulation results show that the proposed model consistently has the minimum statistical errors, and outperforms other benchmark methods.

It can be concluded that the proposed model is conductive to improving the accuracy of multi-step ahead wind speed forecasting.

The novelty and innovation of this study can be described as follows: (a) Unlike the common filter-based forecasting methods which establish forecasting models for each sub-signal resolved by filter algorithms, this study develops an optimal feature extraction algorithm to recombine the optimal features for balancing the calculation efficiency and prediction accuracy; (b) The long-term and short-term memory natures of wind speed change is often neglected in the most existing research. In order to predict wind speed more precisely, a special kind of RNN called as LSTMN is utilized to extract the deep inherent features and detect the long-term and short-term memory natures of wind speed change; (c) In the common filter-based forecasting methods, the analysis of the error component is often neglected because of the hypothesis of white noise, but this may cause a big confusion if the inherent correlation and heteroscedasticity exist in error component. In this study, an ECS based on the hybrid of LSTMN and GARCH is developed to enhance the prediction precision of wind speed. (d) Considering the upper issues, this study develops a novel hybrid prediction model that integrates the merits of OFE, DLA and ECS to strengthen the prediction precision of wind speed.

This paper is organized as follows. Section 2 describes the basic methods used in the proposed model. Section 3 shows the framework of the developed model. Section 4 shows the experimental results and discussion, and Section 5 summarizes the conclusion.

2. Related methodology

This study develops a novel hybrid prediction model that integrates the merits of OFE, DLA and ECS to enhance the prediction accuracy of wind speed. In this section, the basic methods used in the proposed model are introduced as follows.

2.1. VMD algorithm

As a novel filter algorithm, the VMD is often employed to decompose a complicated signal into several quasi-orthogonal intrinsic mode functions [30]. In this study, the VMD is introduced to conduct the decomposition of wind speed signal. Suppose that $\{x(t)\}$ is a wind speed signal, $\sigma(t)$ is the Dirac distribution, μ_k is the k-th mode and ω_k is the corresponding center frequency of the mode ω_k. The main steps of the VMD are demonstrated as follows [31]:

Step 1: Calculate the analytic signal of each mode by Hilbert transform:

$$\left(\sigma(t)+\frac{j}{\pi t}\right)*\mu_k(t) \tag{1}$$

Step 2: Use the corresponding center frequencies to modulate the analytic signal:

$$\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mu_k(t)\right]e^{-j\omega_k t} \tag{2}$$

Step 3: Convert the process of VMD into the constrained optimization problem:

$$\min_{\{\mu_k\},\{\omega_k\}}\left\{\sum_k \left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mu_k(t)\right]e^{-j\omega_k t}\right\|\right\} \quad s.t.\ \sum_k \mu_k = x \tag{3}$$

Step 4: Convert the above constrained problem into the unconstrained problem by introducing the quadratic penalty α and the Lagrangian multipliers $\lambda(t)$:

$$L\{\{\mu_k\},\{\omega_k\},\lambda\} = \alpha\sum_k \left\|\partial_t\left[\left(\sigma(t)+\frac{j}{\pi t}\right)*\mu_k(t)\right]e^{-j\omega_k t}\right\|^2 + \|x(t) - \sum_k \mu_k(t)\|_2^2 + \left\langle \lambda(t), \left(x(t) - \sum_k \mu_k(t)\right)\right\rangle \tag{4}$$

Step 5: Employ the alternate direction method of multipliers (ADMM) to update μ_k^{n+1}, ω_k^{n+1} and λ^{n+1} in two directions, and solve the above unconstrained problem:

$$\hat{\mu}_k^{n+1}(\omega) = \frac{\hat{x}(\omega) - \sum_{i\neq k}\hat{\mu}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega-\omega_k)^2} \tag{5}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega|\hat{\mu}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{\mu}_k(\omega)|^2 d\omega} \tag{6}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau\left(\hat{x}(\omega) - \sum_{k=1}^K \hat{\mu}_k^{n+1}(\omega)\right) \tag{7}$$

where n denotes the number of iterations, τ denotes the update parameter, $\hat{x}(\omega)$, $\hat{\mu}_i(\omega)$, $\hat{\lambda}(\omega)$ and $\hat{\mu}_i^{n+1}(\omega)$ represent the Fourier transforms of $x(t)$, $\mu_i(t)$, $\lambda(t)$ and $\mu_i^{n+1}(t)$, respectively.

2.2. Two feature selection algorithms

2.2.1. KLD-based feature selection

In this study, the KLD is employed to extract meaningful features and remove the disturbance of illusive components from these subseries decomposed by VMD. Given $f(x)$ and $g(\mu_k)$ are the probability density functions (PDFs) of the original wind speed signal $\{x(t)\}$ and the kth subseries $\{\mu_k(t)\}$ decomposed by VMD, respectively. The KLD value between them can be calculated as follows [32]:

$$KLD(f\|g) = \sum_{x\in\{x(t)\}} f(x)\log\frac{f(x)}{g(x)} \tag{8}$$

$$KLD(g\|f) = \sum_{\mu_k\in\{\mu_k(t)\}} g(\mu_k)\log\frac{g(\mu_k)}{f(\mu_k)} \tag{9}$$

where the KLD meets two conditions: (a) the larger the value of KLD, the larger the difference between two signals and vice versa; (b) if the value of KLD is equal to zero, it indicates that two signals are the same.

This study adopts a symmetric version of KLD to conduct the feature extraction process, which is defined as:

$$KLD = \frac{1}{2}\left(\sum_{x\in\{x(t)\}} f(x)\log\frac{f(x)}{g(x)} + \sum_{\mu_k\in\{\mu_k(t)\}} g(\mu_k)\log\frac{g(\mu_k)}{f(\mu_k)}\right) \tag{10}$$

In order to determine $f(x)$ and $g(\mu_k)$, we adopt a non-parametric kernel smoothing approach to estimate them. Take $f(x)$ estimation as an example. $f(x)$ can be estimated as:

$$f(x) \approx \frac{1}{nh}\sum_{t=1}^n L\left(\frac{x-x(t)}{h}\right) \tag{11}$$

where h and $L(\cdot)$ represent the smoothing parameter and symmetric kernel function, respectively, which are given as follows:

$$h = \left(\frac{4\sigma^5}{3n}\right)^{1/5} \tag{12}$$

$$L(v) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{v^2}{2}\right) \tag{13}$$

where σ denotes the sample standard deviation.

We can obtain the $f(x)$ by substituting Eqs. (12) and (13) into Eq. (11). Similarly, we can obtain the $g(\mu_k)$. Therefore, we can obtain the symmetric KLD by Eq. (10). Further, we eliminate the subseries with the maximum KLD value.

2.2.2. EM-based feature selection

In this study, the EM of signal strength is utilized to further detect the illusive components ignored by KLD. The EM value of each sub-signal can be calculated as follows [33]:

$$E_{\mu_k(t)} = \sum_t [\mu_k(t)]^2 \tag{14}$$

Here, the sub-signal with the least EM value is removed as the illusive component.

2.3. SE algorithm

In this section, in order to balance the calculation efficiency and prediction accuracy, a novel approximate entropy called as sample entropy (SE) is adopted to recombine the optimal features according to the SE values of the remaining sub-signals by the two feature selection criteria. In general, the larger of the SE values, the lower sequence autocorrelation are and vice versa. Suppose that N and σ are the length and standard deviation of time series $\{x(t)\}$, r and m represent the similarity tolerance and the dimension, respectively. The SE value $SE(N, m, r)$ can be calculated as follows [34]:

Step 1: Convert the time series $\{x(t)\} = \{x(1), x(2), ...,x(N)\}$ into m-dimensional vector:

$$X_t = [x(t), x(t+1), ...,x(t+m-1)], \quad (t = 1, 2, ...,N-m+1) \tag{15}$$

Step 2: Define $D_m(X_t, X_s)$ as the maximum difference of corresponding elements between the vector X_t and X_s, and calculate each $D_m(X_t, X_s)$, where $t = 1, 2, ...,N-m+1, t \neq s$.

$$D_m(X_t, X_s) = \max_{0\sim m-1}|x_{t+i}-x_{s+i}| \tag{16}$$

Step 3: Define $B_t^m(r)$ as the sum number of $D_m(X_t, X_s) < r$ and obtain the mean value $B^m(r)$ of $B_t^m(r)$:

$$B_t^m(r) = \frac{1}{N-m} num\{D_m(X_t, X_s) < r\} \tag{17}$$

$$B^m(r) = \frac{1}{N-M+1} \sum_{t=1}^{N-M+1} B_t^m(r) \tag{18}$$

Step 4: Replace the m as $m+1$, and obtain the mean value $B^{m+1}(r)$ of $B_t^{m+1}(r)$ by repeating the step 1–3.

$$B^{m+1}(r) = \frac{1}{N-M} \sum_{t=1}^{N-M} B_t^m(r) \tag{19}$$

Step 5: Calculate the SE value by the following Eqs. (20) and (21).

$$SE(m, r) = \lim_{N\to\infty}\left\{-\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right)\right\} \quad (N \text{ is an infinite value}) \tag{20}$$

$$SE(N, m, r) = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right) \quad (N \text{ is a finite value}) \tag{21}$$

Generally, the performance of SE depends on the parameters r and m. In this study, the similarity tolerance r and the dimension m are set to 0.2σ and 2, respectively.

2.4. LSTMN model

As a novel deep learning model, the LSTMN with the memory cell can add or remove information to the memory cell by three controlling gates including input gate, forget gate and output gate. The input gate is used to determine whether the new input information is added to the cell. The forget gate is employed to determine whether the past cell status is removed from the cell. The output gate is utilized to determine whether the latest cell output is propagated to the ultimate state. The structure of the LSTM model can be illustrated in Fig. 1. In this study, the LSTMN is applied to detect the long-term and short-term memory natures of wind speed change.

Suppose that x_t is the actual wind speed value at time t, $\hat{x}_t$ is the corresponding prediction value of wind speed x_t, i_t represents the input gate, f_t and o_t represent the forget gate and output gate, respectively. The main process of the LSTMN model can be given as follows [35]:

$$i_t = \sigma(W_{ix}x_t + W_{ip}p_{t-1} + W_{ic}c_{t-1} + b_i) \tag{22}$$

$$f_t = \sigma(W_{fx}x_t + W_{fp}p_{t-1} + W_{fc}c_{t-1} + b_f) \tag{23}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g(W_{cx}x_t + W_{cp}p_{t-1} + b_c) \tag{24}$$

$$o_t = \sigma(W_{ox}x_t + W_{op}p_{t-1} + W_{oc}c_t + b_o) \tag{25}$$

$$y_t = o_t \circ z(c_t) \tag{26}$$

$$\hat{x}_t = W_{xy}y_t + b_x \tag{27}$$

where '$\circ$' represents the scalar product, W represents the weight matrices, b represents the bias vectors, c_t represents the activation vector of each cell, y_t represents the activation vector of each memory block, $\sigma(\cdot)$ represents the standard logistic function ($\sigma(x) = \frac{1}{1+e^{-x}}$), $g(\cdot)$ represents the centered logistic function ($g(x) = \frac{4}{1+e^{-x}} - 2\ x \in [-2, 2]$), and $z(\cdot)$ represents the centered logistic function ($z(x) = \frac{2}{1+e^{-x}} - 1\ x \in [-1, 1]$).

2.5. GARCH model

In this section, the GARCH model is employed to verify the volatility of the errors and correct the prediction results of LSTMN model. Given x_t and $\hat{x}_{LSTMN}(t)$ are the actual wind speed value at time t and the corresponding prediction value of LSTMN model, respectively. The actual value x_t can be described as follows [36]:

$$x_t = \hat{x}_{LSTMN}(t) + \xi_t \tag{28}$$

where ξ_t denotes the prediction error of the LSTMN model at time t. If ξ_t has the time-varying variance, it is given as follows:

$$\xi_t = \sqrt{\delta_t}\varepsilon_t \tag{29}$$

where ε_t represents the white noise sequence; δ_t represents the conditional variance and is given as follows:

$$\delta_t = \varsigma_0 + \sum_{k=1}^{p} \varsigma_k \delta_{t-k} + \sum_{l=1}^{q} \eta_l \xi_{t-l}^2$$
$$\begin{cases} \varsigma_0 > 0 \\ \varsigma_k \geqslant 0\ k = 1, 2, ..., p \\ \eta_l \geqslant 0\ l = 1, 2, ..., q \\ \sum_{k=1}^{p} \varsigma_k + \sum_{l=1}^{q} \eta_l < 1 \end{cases} \tag{30}$$

where η_l and ς_k are nonnegative varying coefficients; p and q denote the orders, respectively. Thus, the error ξ_t follows $GARCH(p, q)$ model.

The Lagrange Multiplier (LM) is usually applied to determine whether it is essential to construct the GARCH model. The LM statistic is given as follows [37]:

$$LM = nr^2 \sim \chi^2(q) \tag{31}$$

where r^2 denotes the goodness of fit. If $LM\ value > \chi^2(q)$, the GARCH is employed to modify the prediction results of LSTMN model. In this study, the standard GARCH (1, 1) is utilized to correct the prediction results of LSTMN model.

3. Proposed model

As presented in introduction, this study develops a novel forecasting model to strengthen the prediction precision of wind speed. This section gives the main procedure of the developed model. Fig. 1 demonstrates the flowchart of the developed model and the whole process is introduced as follows:

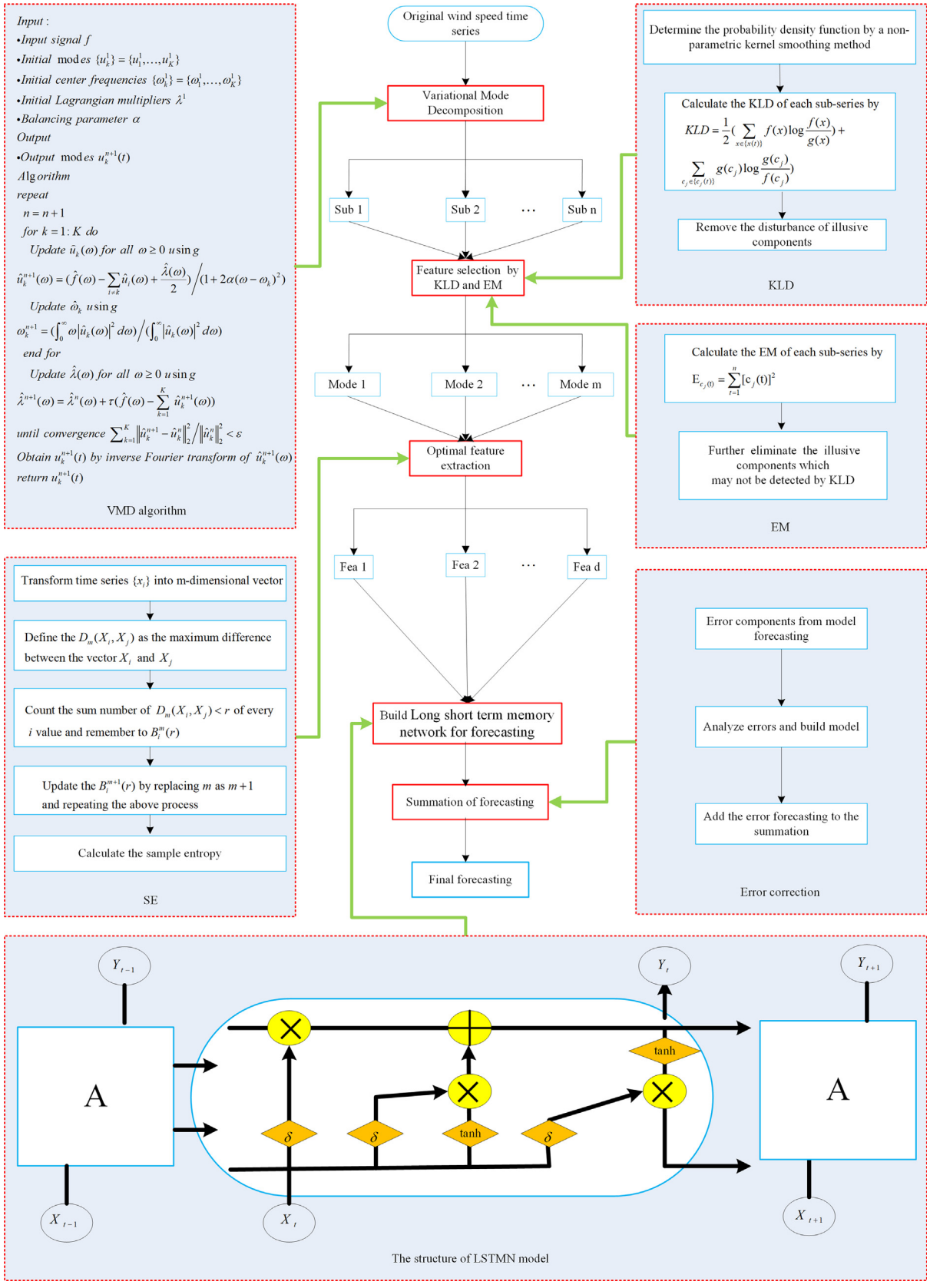


Fig. 1. Flowchart of the proposed model.

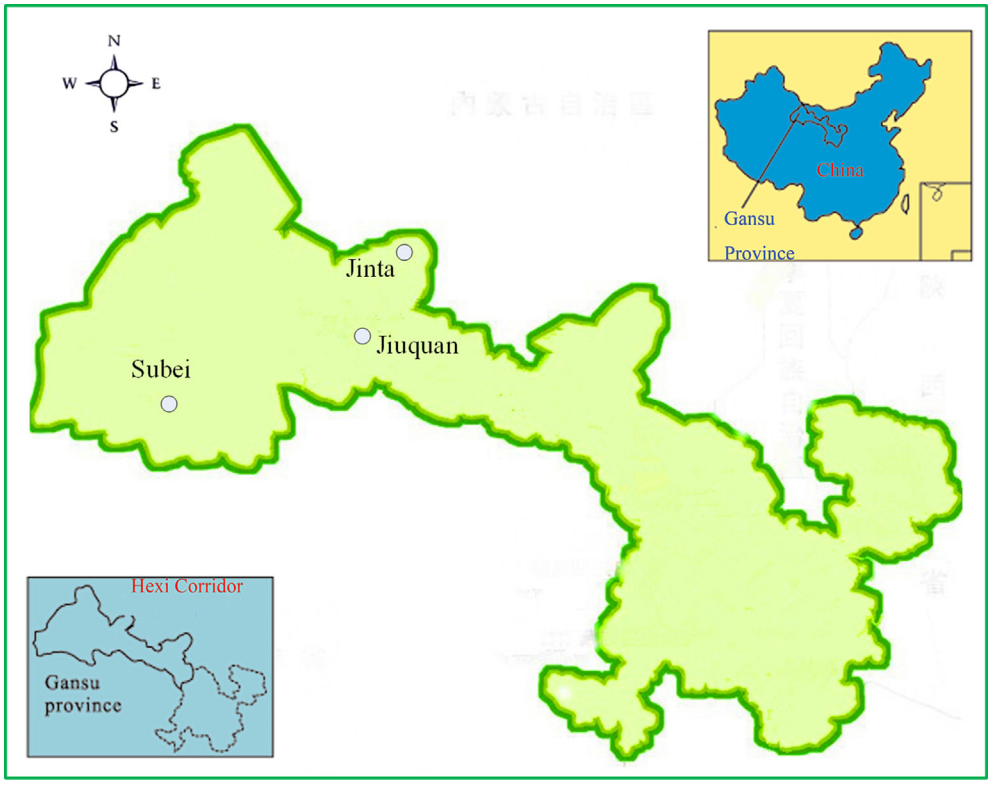


Fig. 2. Three wind farms considered for the experiments.

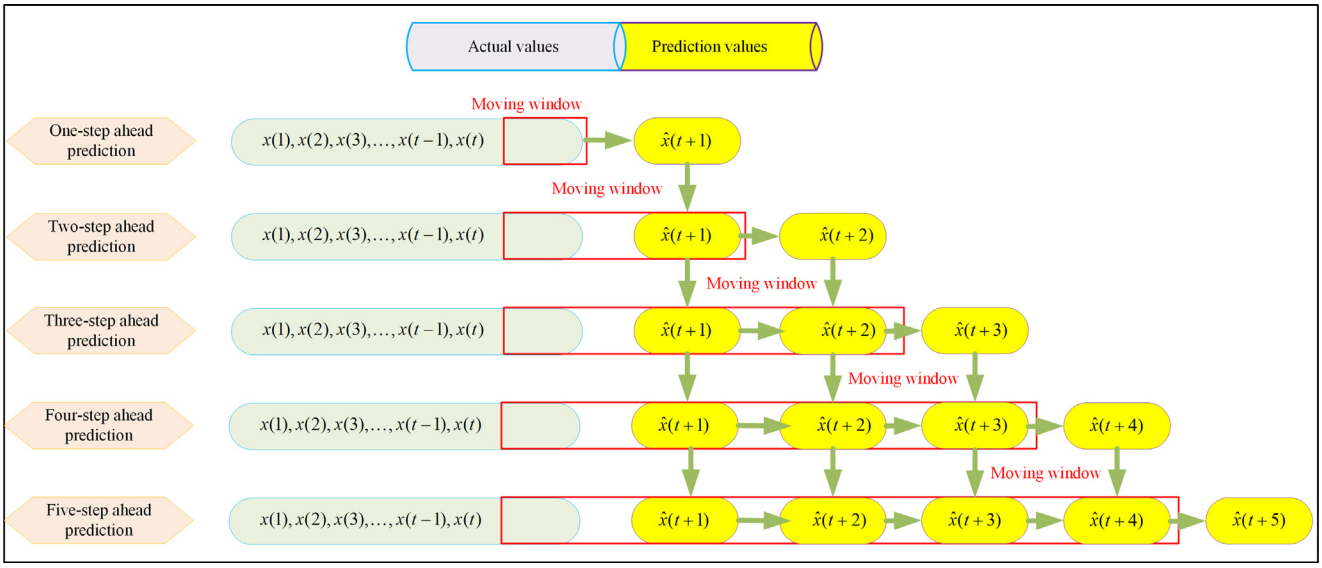


Fig. 3. The architecture of multi-step ahead prediction.

Table 1
Three statistical measures used in the analysis of prediction results.

Error criteria	Definition	Equation
MAE	Mean absolute error	$\frac{1}{n_1}\sum_{t=1}^{n_1}\lvert x(t)-\hat{x}(t)\rvert$
RMSE	Root mean square error	$\sqrt{\frac{1}{n_1}\sum_{t=1}^{n_1}(x(t)-\hat{x}(t))^2}$
MAPE	Mean absolute percentage error	$\frac{1}{n_1}\sum_{t=1}^{n_1}\lvert\frac{x(t)-\hat{x}(t)}{x(t)}\rvert\times 100\%$

where n_1 denotes the number of prediction samples, $x(t)$ is the actual wind speed value at time t and $\hat{x}(t)$ is the prediction value of wind speed at the same time point.

(1) The VMD is employed to resolve a multi-component wind velocity signal into n sub-signals.
(2) Two feature selection algorithms including KLD and EM are applied to eliminate the illusive subseries and reserve m useful subseries from all subseries decomposed by VMD, where the m is smaller than the number n, achieving the goal of getting the initial features from the original wind speed signal.
(3) The SE algorithm is utilized to recombine the m initial features subseries and obtain d optimal features subseries by calculating their SE values, where d is smaller than the number m, achieving the goal of reducing the workload and saving the computation time for prediction.

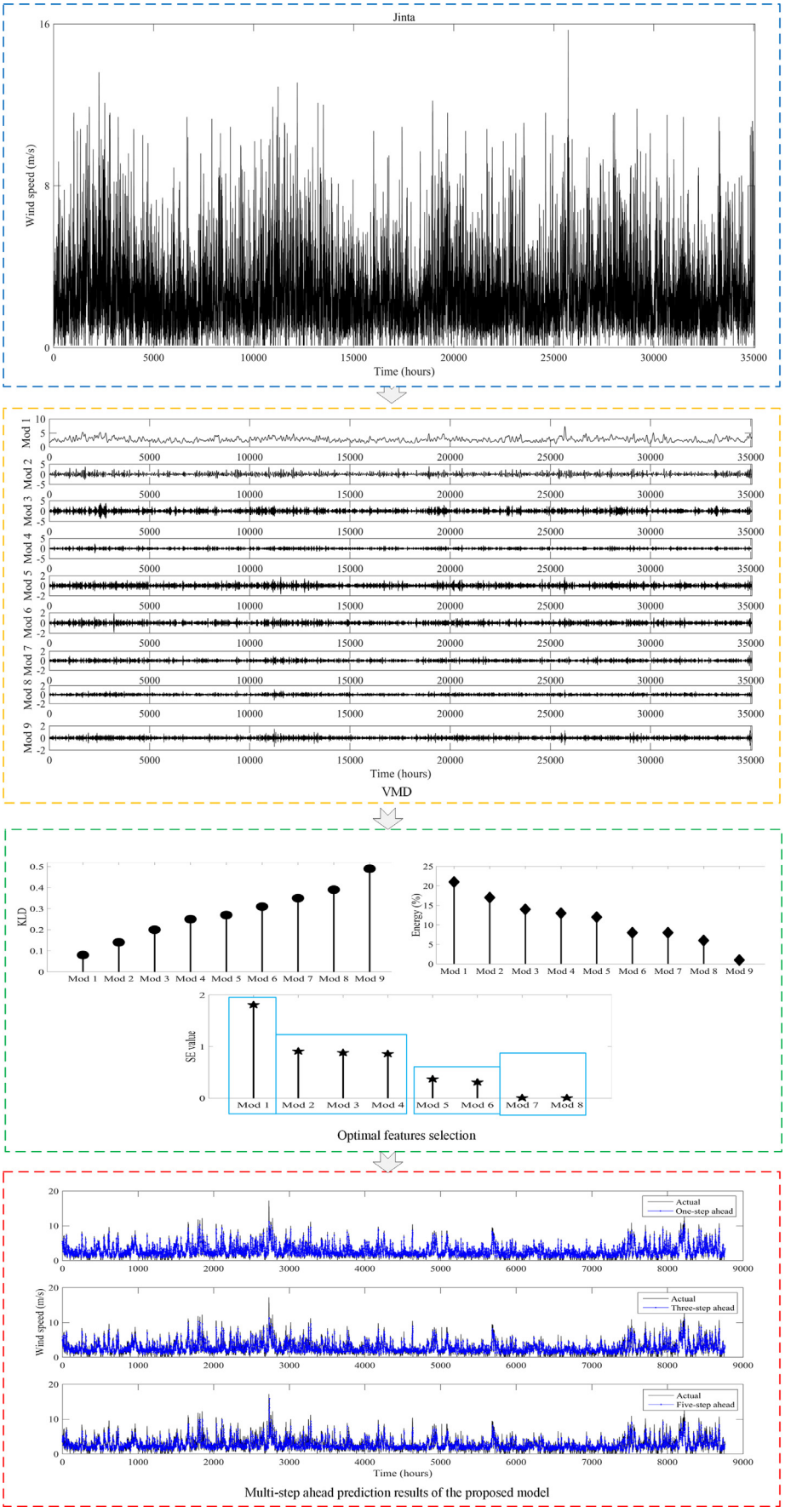


Fig. 4. The prediction process of the proposed model in Jinta.

Table 2
These benchmark models for comparisons.

Model 1	SVR
Model 2	ELM
Model 3	LSTMN
Model 4	VMD-LSTMN
Model 5	VMD-OFE-LSTMN

(4) The LSTMN model is built to forecasting these optimal feature subseries.
(5) An ECS is presented to strengthen the forecasting precision of wind speed. Both autocorrelation (ACF) and partial autocorrelation (PACF) are firstly used to analyze the interdependence structure of the error components and determine whether it is essential to construct the corresponding prediction model for them. Then, the LM is applied to examine the heteroscedasticity of error components and determine whether it is essential to construct the GARCH model. In view of the above analysis, there are four different ways of handling the error components: (i) the hybrid LSTMN-GARCH model is employed to deal with the error components when there exists both characteristics of correlation and heteroscedasticity in error components; (ii) the LSTMN model is applied to modify the error components when there exists only the correlation in error components; (iii) the GARCH model is adopted to modify the error components when there exists only the heteroscedasticity in error components; (iv) if both characteristics of correlation and heteroscedasticity are not obvious in error components, it is not essential to modify the error components.

4. Case study

4.1. Data collection

The Hexi corridor of China called as “wind corridor”, has abundant wind resource because of its special geographical location. In this study, three large hourly wind speed datasets collected from three wind farms including Jinta, Subei and Jiuquan lying in the Hexi corridor, are adopted to compare the prediction precision of the developed model with other benchmark models. Fig. 2 displays three wind farms considered for the experiments. Each dataset includes 43,824 samples. The first 35,064 samples of each wind speed dataset are utilized to establish the model, and the remaining samples are adopted to verify the prediction ability of the developed model. Moreover, a rolling prediction mechanism is employed to conduct the multi-step prediction. Fig. 3 shows the architecture of multi-step ahead prediction.

4.2. Evaluation criteria

In this study, three statistical measures including MAE, RMSE and MAPE are adopted to quantitatively assess the forecasting ability of all involved models, which are listed in Table 1. Generally, the smaller values of three statistical measures values, the better prediction ability of the corresponding model is and vice versa.

4.3. Experiment I: test with wind speed from case 1

This section adopts a wind speed dataset measured from Jinta wind farm to verify the prediction ability of the proposed model thoroughly. First, the VMD is utilized to resolve the original wind speed into nine sub-signals including Mod 1, Mod 2, Mod 3, Mod 4, Mod 5, Mod 6, Mod 7, Mod 8 and Mod 9. Then, two feature selection algorithms including KLD and EM are employed to extract meaningful features from these sub-signals and remove the disturbance of illusive sub-signals. According to the above two feature selection criteria, we obtain eight initial features including Mod 1, Mod 2, Mod 3, Mod 4, Mod 5, Mod 6, Mod 7 and Mod 8, and eliminate the Mod 9 as noise. Further, in order to balance the calculation efficiency and prediction accuracy of the model, the SE algorithm is utilized to recombine the optimal features sub-signals from these initial features. Four optimal features sub-signals including Fea1 (Mod 1), Fea2 (Mod 2, Mod 3 and Mod 4), Fea3 (Mod 5 and Mod 6) and Fea4 (Mod 7, and Mod 8) are recombined by calculating their SE values. For each feature sub-signal, the LSTMN is used to establish the corresponding prediction model. The ECS is adopted to further improve the prediction precision of each feature sub-signal. Both ACF and PACF are firstly used to analyze the interdependence structure of the error components between the actual values of wind speed and prediction values of the LSTMN model, and determine whether it is essential to construct the corresponding prediction model for the error components. Then, the LM is applied to examine whether it is essential to construct the GARCH. In view of the above analysis, we obtain the prediction values of each feature sub-signal. Further, we obtain the final forecasting results of raw wind speed by aggregating the forecasting values of each feature sub-signal. Fig. 4 presents the forecasting results of the developed model in Jinta. As seen from Fig. 4, the proposed model can obviously capture the main trends of wind speed change.

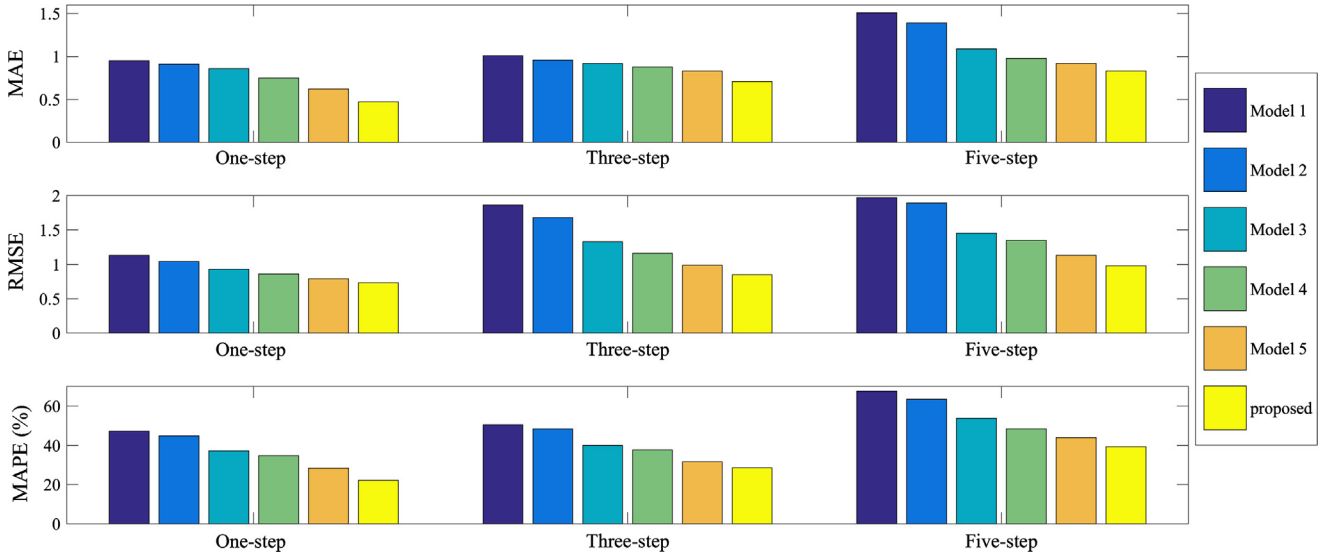


Fig. 5. Prediction errors comparison of different models in Jinta.

Table 3
Three errors comparison of different models in Jinta (MAE (m/s), RMSE (m/s), MAPE (%)).

Models	One-step ahead			Three-step ahead			Five-step ahead		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
SVR	0.95	1.13	47.16	1.01	1.86	50.43	1.51	1.97	67.58
ELM	0.91	1.04	44.71	0.96	1.68	48.42	1.39	1.89	63.45
LSTMN	0.86	0.93	37.14	0.92	1.33	39.94	1.09	1.45	53.71
VMD-LSTMN	0.75	0.86	34.81	0.88	1.16	37.65	0.98	1.35	48.32
VMD-OFE-LSTMN	0.62	0.79	28.33	0.83	0.99	31.64	0.92	1.13	43.87
VMD-OFE-LSTMN-ECS	0.47	0.73	22.16	0.71	0.85	28.59	0.83	0.98	39.32

To further verify the prediction ability of the developed model, five benchmark models are involved in the comparison of the prediction performance and are listed in Table 2. Multi-step ahead prediction is conducted and three corresponding error results are shown in Fig. 5 and Table 3. As seen from Fig. 5 and Table 3, it can be found as follows:

(1) Compared with Model 1 and Model 2, Model 3 has the smallest errors for multi-step ahead forecasting including one-step, three-step and five-step. For instance, the MAPEs of Model 3 are 37.14%, 39.94% and 53.71% in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding MAPEs of Model 1 are 47.16%, 50.43% and 67.58%. The proposed model reflects 10.02%, 10.49% and 13.87% improvements, respectively. This is mainly put down to the fact that the LSTMN model can capture the deep features in wind speed very well.
(2) Compared with Model 4, Model 5 has the better performance for multi-step ahead prediction. For instance, the MAEs of Model 5 are 0.62, 0.83 and 0.92 in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding MAEs of Model 4 are 0.75, 0.88 and 0.98. This indicates that the OFE is conducive to raising the prediction accuracy of wind speed.
(3) The developed model has the better performance than the Model 5. For instance, the RMSEs of the developed model are 0.73, 0.85 and 0.98 in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding RMSEs of Model 5 are 0.79, 0.99 and 1.13. This can be attributed to the fact that the ECS is effective in enhancing the forecasting ability.
(4) The developed model has the best prediction performance than other five benchmark models in horizons of 1-step prediction, 3-step prediction and 5-step prediction. This can be attributed to the fact that the combination of OFE, DLA and ECS is help for enhancing the forecasting accuracy.

4.4. Experiment II: test with wind speed from case 2

The experiment II adopts a wind speed dataset measured from Subei wind farm to validate the developed model. Fig. 6 indicates the prediction process of the developed model in Subei farm. Fig. 7 and Table 4 present the forecasting results of different models and the corresponding estimated errors results including MAE, RMSE and MAPE. As seen from Fig. 7 and Table 4, it can be concluded that: (1) the LSTMN model has precedence to the traditional SVR and ELM models in multi-step ahead forecasting, which indicates that the LSTMN model can detect the long-term and short-term memory natures of wind speed change and strengthen the ability for extracting the deep inherent features of wind speed; (2) the optimal features extraction algorithm affords a positive effect in strengthening the prediction performance of the VMD-LSTMN model; (3) the ECS is conducive to enhancing the forecasting precision of wind speed; (4) Compared with other benchmark models, the developed model consistently has the minimum statistical errors and is more effective for multi-step ahead wind speed prediction.

4.5. Experiment III: test with wind speed from case 3

To further verify the stability of the developed model, the experiment III is conducted by the wind speed data from Jiuquan farm. Fig. 8 indicates the prediction process of the proposed model in Jiuquan farm. Fig. 9 and Table 5 indicate the evaluation procedures. The similar conclusions can be concluded from Fig. 9 and Table 5.

4.6. Summary based on experiments I-III

Table 6 indicates the average forecasting ability of different models in three cases. As seen from Table 6, the following conclusions can be summarized:

(1) The LSTMN model, as an effective DLA, can catch the deep inherent characteristics of wind speed change and outperform the traditional SVR and ELM models for multi-step forecasting in three cases.
(2) The OFE algorithm makes a pivotal contribution for multi-step forecasting in three cases. For instance, the mean MAPEs of Model 5 are 28.34%, 34.07% and 41.45% in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding mean MAPEs of Model 4 are 33.19%, 38.53% and 45.10%, which reflects the improvements of 4.85%, 4.46%, 3.65%, respectively. This indicates that the OFE is conducive to raising the precision of multi-step prediction.
(3) The ECS affords a positive effect for multi-step prediction in three cases. For example, the mean MAEs of the proposed model are 0.43, 0.69 and 0.83 in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding mean MAEs of Model 5 are 0.57, 0.79 and 0.93. This indicates that the ECS has the ability to strengthen the performance of multi-step prediction.
(4) The developed model indicates the superior and trustworthy ability than other benchmark models in three cases, and it has great potential for wind speed prediction.

4.7. Analysis and discussion

In this section, OFE and ECS are discussed to uncover their effects on the developed model, respectively.

4.7.1. Discussion of the advantage of the OFE

In order to further verify the advantage of the OFE, the OFE is used to improve these common filter-based models for multi-step prediction. These common filter-based models, mainly including WD-LSTMN-ECS, EMD-LSTMN-ECS, WPD-LSTMN-ECS and EEMD-LSTMN-ECS, are chosen as benchmark models to validate the excellent ability of OFE. Tables 7–9 indicate the errors comparison of different models with feature selection and without feature selection in three cases. As seen from Tables 7–9, it can be concluded as follows:

(1) The WD-OFE-LSTMN-ECS is superior to the WD-LSTMN-ECS for multi-step forecasting in three cases. For example, in Jinta, the

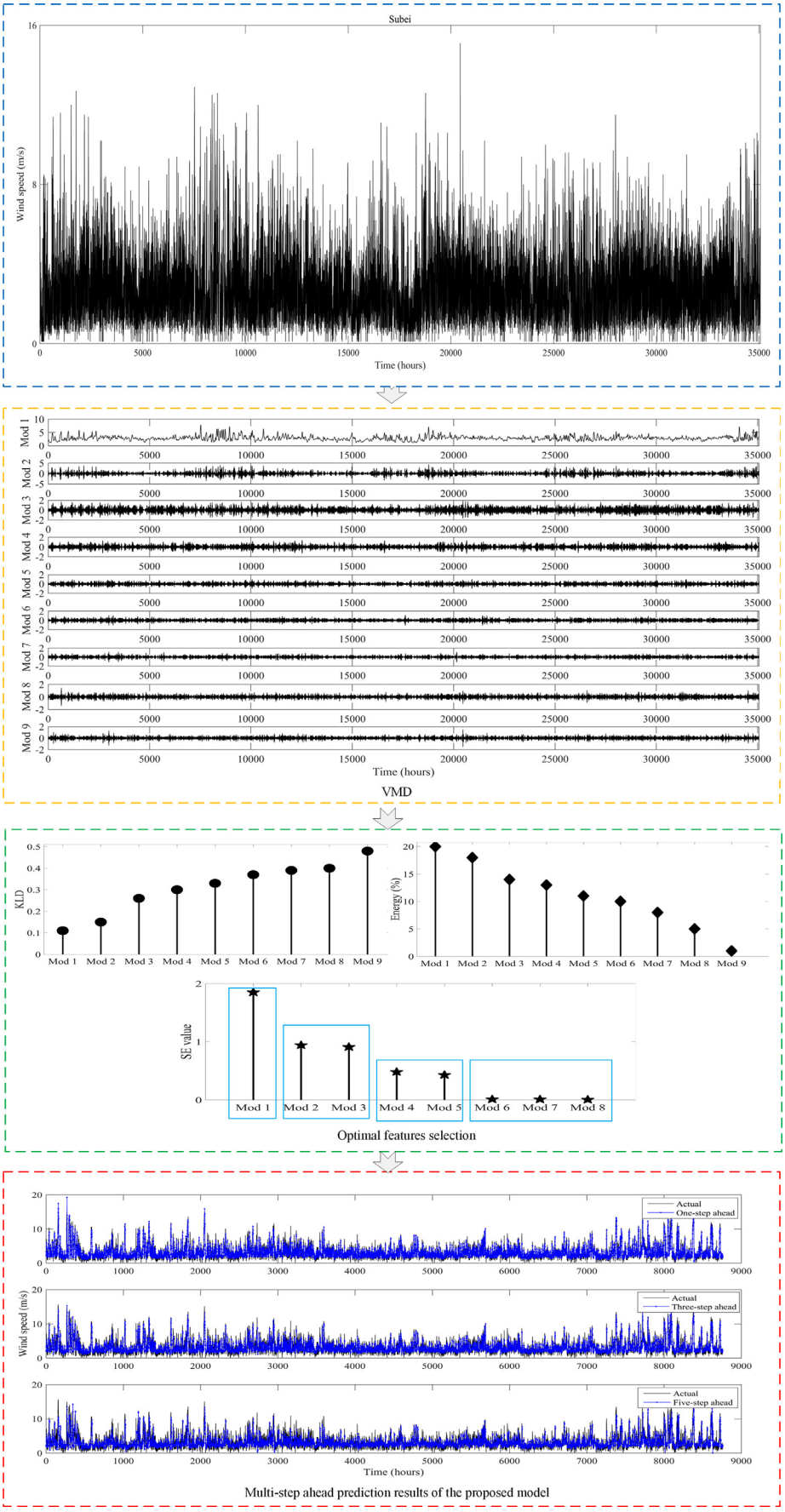


Fig. 6. The prediction process of the proposed model in Subei.

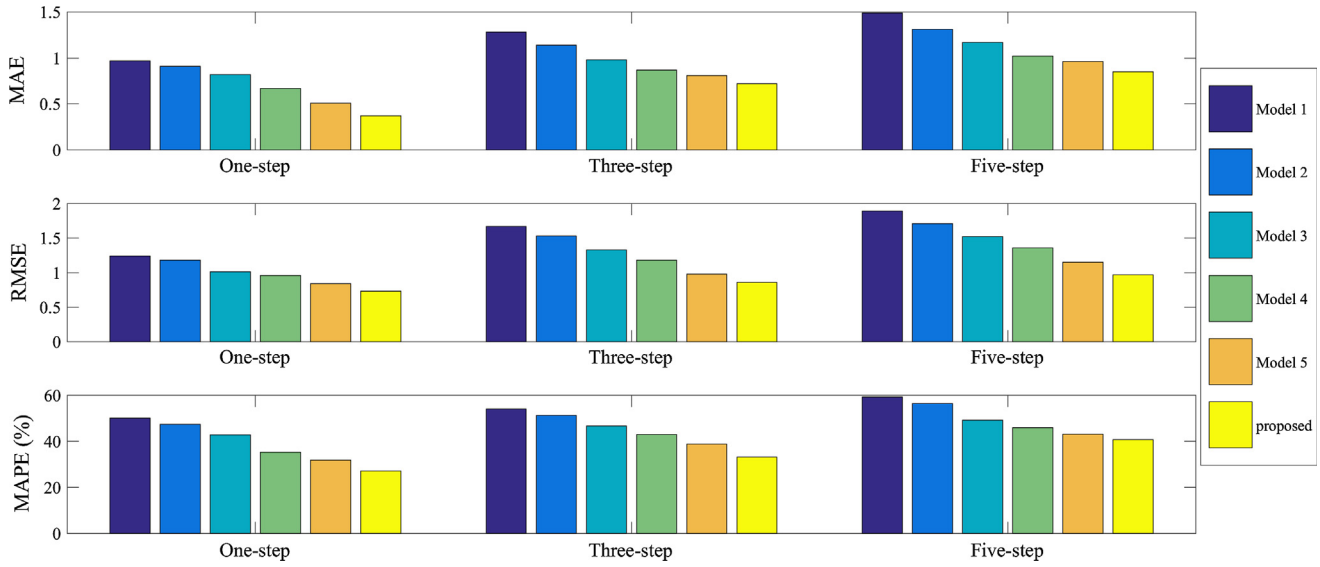


Fig. 7. Prediction errors comparison of different models in Subei.

Table 4
Three errors comparison of different models in Subei (MAE (m/s), RMSE (m/s), MAPE (%)).

Models	One-step ahead			Three-step ahead			Five-step ahead		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
SVR	0.97	1.24	50.13	1.28	1.67	54.04	1.49	1.89	59.23
ELM	0.91	1.18	47.46	1.14	1.53	51.29	1.31	1.71	56.48
LSTMN	0.82	1.01	42.84	0.98	1.33	46.72	1.17	1.52	49.16
VMD-LSTMN	0.67	0.96	35.23	0.87	1.18	42.94	1.02	1.36	45.91
VMD-OFE-LSTMN	0.51	0.84	31.92	0.81	0.98	38.83	0.96	1.15	43.05
VMD-OFE-LSTMN-ECS	0.37	0.73	27.15	0.72	0.86	33.14	0.85	0.97	40.71

MAPEs of the WD-OFE-LSTMN-ECS model are 27.86%, 31.16% and 43.11% in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding MAPEs of the WD-LSTMN-ECS are 33.18%, 35.81% and 46.51%, which reflects the improvements of 5.32%, 4.65%, 3.4%, respectively. This indicates that the OFE is conducive to raising the precision of multi-step prediction.

(2) The EMD-OFE-LSTMN-ECS outperforms the EMD-LSTMN-ECS for multi-step prediction in three cases. For example, in Subei, the MAEs of the EMD-OFE-LSTMN-ECS are 0.48, 0.75 and 0.93 in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding MAEs of the EMD-LSTMN-ECS are 0.63, 0.83 and 0.98, respectively. This indicates that the OFE has the ability to strengthen the accuracy of multi-step prediction.

(3) The WPD-OFE-LSTMN-ECS has better performance than the WPD-LSTMN-ECS in three cases. For example, in Jiuquan, the RMSEs of the WPD-OFE-LSTMN-ECS are 0.78, 0.96 and 1.09 in horizons of one-step prediction, three-step prediction and five-step prediction, while the corresponding RMSEs of the WPD-LSTMN-ECS are 0.90, 1.05 and 1.18, respectively. This indicates that the OFE has the ability to strengthen the precision of multi-step prediction.

(4) Similarly, the EEMD-OFE-LSTMN-ECS has better prediction ability than EEMD-LSTMN-ECS in horizons of one-step prediction, three-step prediction and five-step prediction. In brief, the OFE is an effective way to enhance the forecasting ability of these common filter-based models for multi-step prediction.

4.7.2. Discussion of the advantage of the ECS

In this section, in order to further verify the advantage of the ECS, the ECS is utilized to strengthen the forecasting ability of these OFE-based models for multi-step prediction. These OFE-based models mainly including WD-OFE-LSTMN, EMD-OFE-LSTMN, WPD-OFE-LSTMN and EEMD-OFE-LSTMN, are chosen as benchmark models to validate the excellent ability of ECS. Table 10 indicates the errors comparison of different models with ECS and without ECS in three cases. From Table 10, it can be concluded that: (1) the WD-OFE-LSTMN-ECS model is superior to the WD-OFE-LSTMN model in horizons of one-step prediction, three-step prediction and five-step prediction; (2) The EMD-OFE-LSTMN-ECS model outperforms the EMD-OFE-LSTMN model for multi-step prediction; (3) the WPD-OFE-LSTMN-ECS model has the better forecasting ability than the WPD-OFE-LSTMN model; (4) the EEMD-OFE-LSTMN-ECS model has the smaller errors than the EEMD-OFE-LSTMN model in three cases; (5) in short, the ECS is conducive to raising the multi-step prediction ability of these OFE-based models.

5. Conclusions

Forecasting wind speed accurately is a key task in the planning and operation of wind energy generation in power systems, and its importance increases with the high integration of wind power into the electricity market. This research develops an innovative hybrid approach for wind speed prediction including variational mode decomposition, Kullback-Leibler divergence, energy measure, sample entropy, long short term memory network and Generalized auto-regressive conditionally heteroscedastic model. The variational mode decomposition is employed to resolve a non-stationary wind speed signal into several more stationary sub-signals. Kullback-Leibler divergence and energy measure are applied to extract these meaningful features and remove the disturbance of illusive components from these sub-signals. In order to balance the calculation efficiency and prediction accuracy,

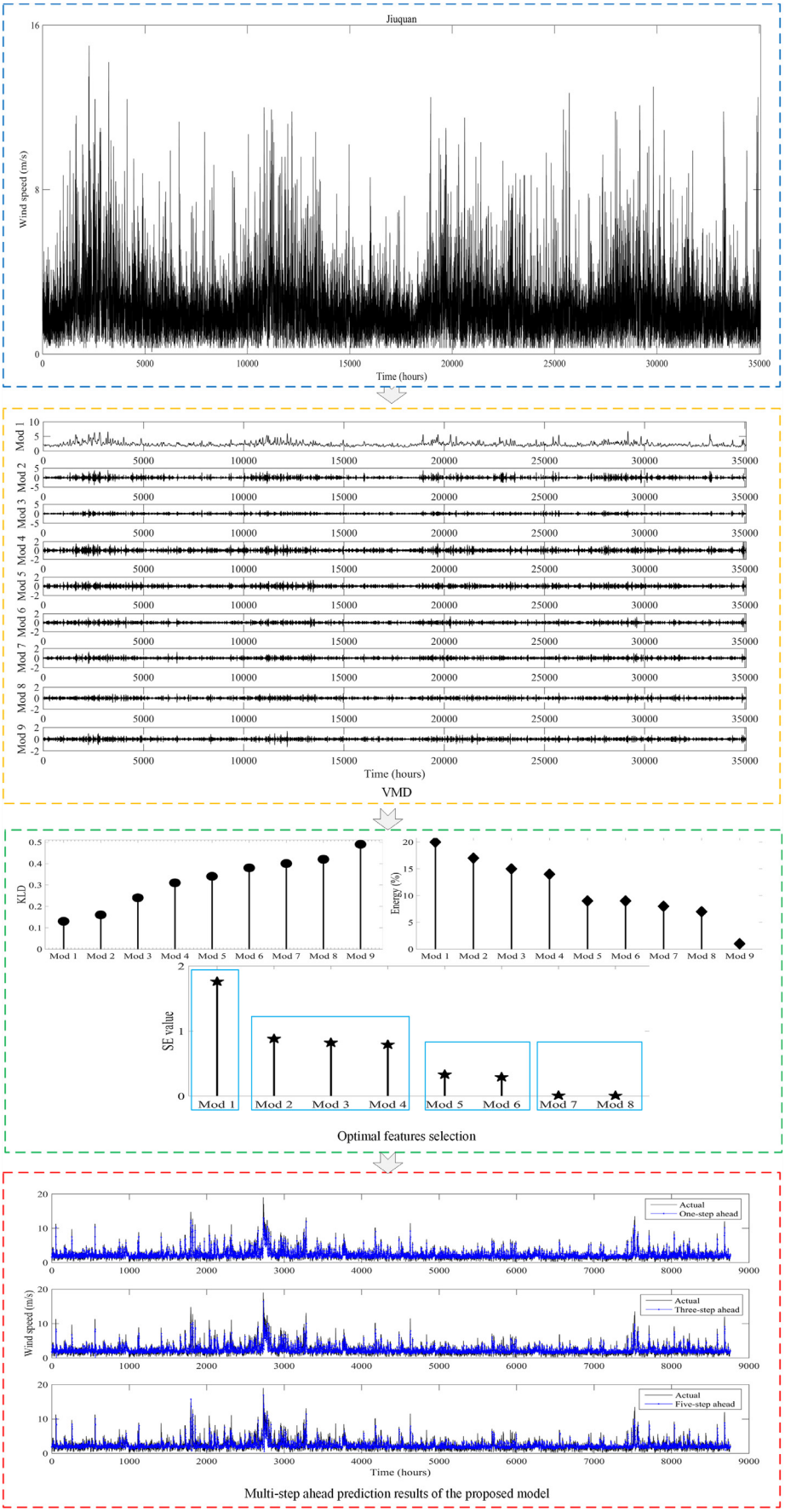


Fig. 8. The prediction process of the proposed model in Jiuquan.

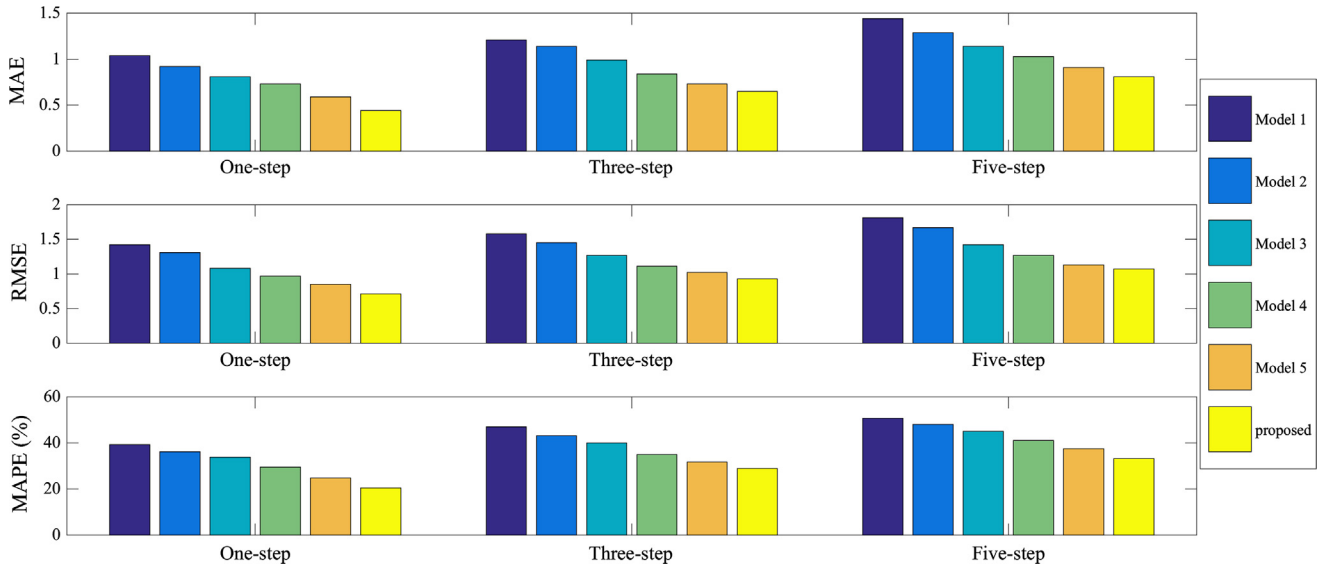


Fig. 9. Prediction errors comparison of different models in Jiuquan.

Table 5
Three errors comparison of different models in Jiuquan (MAE (m/s), RMSE (m/s), MAPE (%)).

Models	One-step ahead			Three-step ahead			Five-step ahead		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
SVR	1.04	1.42	39.25	1.21	1.58	46.92	1.44	1.81	50.62
ELM	0.92	1.31	36.13	1.14	1.45	43.12	1.29	1.67	48.05
LSTMN	0.81	1.08	33.71	0.99	1.27	39.98	1.14	1.42	45.11
VMD-LSTMN	0.73	0.97	29.54	0.84	1.11	35.01	1.03	1.27	41.07
VMD-OFE-LSTMN	0.59	0.85	24.78	0.73	1.02	31.73	0.91	1.13	37.42
VMD-OFE-LSTMN-ECS	0.44	0.71	20.51	0.65	0.93	28.95	0.81	1.07	33.14

Table 6
Average errors comparison of different models in three Cases (MAE (m/s), RMSE (m/s), MAPE (%)).

Models	One-step ahead			Three-step ahead			Five-step ahead		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
SVR	0.99	1.26	45.51	1.17	1.70	50.46	1.48	1.89	59.14
ELM	0.91	1.18	42.77	1.08	1.55	47.61	1.33	1.76	55.99
LSTMN	0.83	1.01	37.90	0.96	1.31	42.21	1.13	1.46	49.33
VMD-LSTMN	0.72	0.93	33.19	0.86	1.15	38.53	1.01	1.33	45.10
VMD-OFE-LSTMN	0.57	0.83	28.34	0.79	1.00	34.07	0.93	1.14	41.45
VMD-OFE-LSTMN-ECS	0.43	0.72	23.27	0.69	0.88	30.23	0.83	1.01	37.72

Table 7
Three errors comparison of different models with feature selection and without feature selection in Jinta (MAE (m/s), RMSE (m/s), MAPE (%)).

Feature selection	Models	One-step ahead			Three-step ahead			Five-step ahead		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Yes	WD-OFE-LSTMN-ECS	0.62	0.78	27.86	0.82	0.98	31.16	0.90	1.11	43.11
	EMD-OFE-LSTMN-ECS	0.59	0.77	26.51	0.81	0.95	30.94	0.88	1.08	42.29
	WPD-OFE-LSTMN-ECS	0.58	0.75	25.01	0.79	0.91	30.07	0.86	1.03	41.42
	EEMD-OFE-LSTMN-ECS	0.49	0.74	23.83	0.78	0.88	29.14	0.85	0.99	40.53
No	WD-LSTMN-ECS	0.72	0.84	33.18	0.86	1.09	35.81	0.97	1.29	46.51
	EMD-LSTMN-ECS	0.69	0.81	30.53	0.85	1.05	32.75	0.95	1.17	45.13
	WPD-LSTMN-ECS	0.67	0.78	29.23	0.84	1.01	32.05	0.93	1.14	44.54
	EEMD-LSTMN-ECS	0.65	0.78	28.91	0.83	0.99	31.74	0.91	1.12	43.98

Table 8
Three errors comparison of different models with feature selection and without feature selection in Subei (MAE (m/s), RMSE (m/s), MAPE (%)).

Feature selection	Models	One-step ahead			Three-step ahead			Five-step ahead		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Yes	WD-OFE-LSTMN-ECS	0.50	0.82	30.23	0.77	0.96	37.14	0.94	1.12	42.87
	EMD-OFE-LSTMN-ECS	0.48	0.80	29.88	0.75	0.93	36.56	0.93	1.10	42.13
	WPD-OFE-LSTMN-ECS	0.45	0.78	28.15	0.74	0.90	35.01	0.90	1.02	41.03
	EEMD-OFE-LSTMN-ECS	0.39	0.75	27.89	0.73	0.87	33.86	0.87	0.99	40.99
No	WD-LSTMN-ECS	0.65	0.94	34.02	0.85	1.15	42.03	0.99	1.31	45.02
	EMD-LSTMN-ECS	0.63	0.91	33.92	0.83	1.11	41.01	0.98	1.28	44.87
	WPD-LSTMN-ECS	0.60	0.89	32.02	0.81	1.06	39.82	0.97	1.21	43.99
	EEMD-LSTMN-ECS	0.53	0.85	30.99	0.80	1.02	38.07	0.95	1.18	43.16

Table 9
Three errors comparison of different models with feature selection and without feature selection in Jiuquan (MAE (m/s), RMSE (m/s), MAPE (%)).

Feature selection	Models	One-step ahead			Three-step ahead			Five-step ahead		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Yes	WD-OFE-LSTMN-ECS	0.57	0.84	24.29	0.72	0.99	31.01	0.90	1.11	37.02
	EMD-OFE-LSTMN-ECS	0.55	0.82	23.04	0.71	0.98	30.76	0.88	1.10	36.11
	WPD-OFE-LSTMN-ECS	0.51	0.78	22.49	0.70	0.96	30.17	0.86	1.09	35.33
	EEMD-OFE-LSTMN-ECS	0.46	0.75	21.01	0.69	0.95	29.23	0.84	1.08	34.64
No	WD-LSTMN-ECS	0.71	0.96	29.14	0.82	1.09	34.51	1.02	1.25	40.01
	EMD-LSTMN-ECS	0.69	0.93	28.06	0.79	1.07	33.35	0.99	1.20	39.42
	WPD-LSTMN-ECS	0.65	0.90	27.11	0.77	1.05	32.94	0.95	1.18	38.75
	EEMD-LSTMN-ECS	0.60	0.86	25.92	0.75	1.03	32.19	0.92	1.15	37.99

Table 10
Three errors comparison of different models with ECS and without ECS in three cases (MAE (m/s), RMSE (m/s), MAPE (%)).

Cases	Models	One-step ahead			Three-step ahead			Five-step ahead		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Jinta	WD-OFE-LSTMN	0.73	0.85	33.12	0.87	1.12	36.01	0.98	1.28	47.04
	EMD-OFE-LSTMN	0.71	0.84	32.45	0.86	1.08	34.15	0.97	1.21	46.67
	WPD-OFE-LSTMN	0.68	0.83	30.98	0.85	1.05	33.06	0.96	1.17	45.82
	EEMD-OFE-LSTMN	0.65	0.82	29.02	0.84	0.99	32.38	0.95	1.15	44.16
	WD-OFE-LSTMN-ECS	0.62	0.78	27.86	0.82	0.98	31.16	0.90	1.11	43.11
	EMD-OFE-LSTMN-ECS	0.59	0.77	26.51	0.81	0.95	30.94	0.88	1.08	42.29
	WPD-OFE-LSTMN-ECS	0.58	0.75	25.01	0.79	0.91	30.07	0.86	1.03	41.42
	EEMD-OFE-LSTMN-ECS	0.49	0.74	23.83	0.78	0.88	29.14	0.85	0.99	40.53
Subei	WD-OFE-LSTMN	0.65	0.94	34.91	0.85	1.15	41.73	1.01	1.33	45.12
	EMD-OFE-LSTMN	0.60	0.91	34.09	0.85	1.11	41.05	0.99	1.28	44.99
	WPD-OFE-LSTMN	0.57	0.89	33.11	0.84	1.06	40.11	0.98	1.22	44.08
	EEMD-OFE-LSTMN	0.52	0.85	32.28	0.83	0.99	39.04	0.97	1.16	43.94
	WD-OFE-LSTMN-ECS	0.50	0.82	30.23	0.77	0.96	37.14	0.94	1.12	42.87
	EMD-OFE-LSTMN-ECS	0.48	0.80	29.88	0.75	0.93	36.56	0.93	1.10	42.13
	WPD-OFE-LSTMN-ECS	0.45	0.78	28.15	0.74	0.90	35.01	0.90	1.02	41.03
	EEMD-OFE-LSTMN-ECS	0.39	0.75	27.89	0.73	0.87	33.86	0.87	0.99	40.99
Jiuquan	WD-OFE-LSTMN	0.71	0.95	28.02	0.81	1.10	34.87	1.01	1.25	40.27
	EMD-OFE-LSTMN	0.68	0.92	27.19	0.78	1.07	34.01	0.99	1.22	39.87
	WPD-OFE-LSTMN	0.61	0.89	26.53	0.76	1.05	33.92	0.97	1.19	38.56
	EEMD-OFE-LSTMN	0.58	0.87	25.29	0.74	1.03	33.86	0.95	1.14	38.02
	WD-OFE-LSTMN-ECS	0.57	0.84	24.29	0.72	0.99	31.01	0.90	1.11	37.02
	EMD-OFE-LSTMN-ECS	0.55	0.82	23.04	0.71	0.98	30.76	0.88	1.10	36.11
	WPD-OFE-LSTMN-ECS	0.51	0.78	22.49	0.70	0.96	30.17	0.86	1.09	35.33
	EEMD-OFE-LSTMN-ECS	0.46	0.75	21.01	0.69	0.95	29.23	0.84	1.08	34.64

the sample entropy is adopted to recombine the optimal features according to the sample entropy value of each sub-signal. Considering the long-term and short-term memory characteristics of wind speed change, the long short term memory network is used to build the prediction model for the optimal feature sub-signals. The error correction strategy based on Generalized auto-regressive conditionally heteroscedastic model is developed to correct the above prediction errors when its inherent correlation and heteroscedasticity cannot be ignored. Three real forecasting cases are applied to verify the performance of the developed approach. The simulation results demonstrate that: (a) the optimal feature extraction is conductive to strengthening the forecasting precision of wind speed with noise; (b) the long short term memory network has a perfect performance in capturing the long-term and short-term memory characteristics of wind speed fluctuations; (c) the error correction strategy is helpful for strengthening the forecasting ability of the model; (d) the proposed model consistently has the minimum statistical error and outperforms other benchmark models; (e) the developed model is suitable for wind speed prediction.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 71501101), the Natural Science Foundation of Jiangsu Province (Grant No. BK20150928), the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (Grant No. 2015SJB063), the China Postdoctoral Science Foundation (Grant No. 2016M601139), the Qing Lan Project, the National Natural Science Foundation of China (Grant No. 11801276, 71603128, 61502242 and 91546117), the National Social Science Foundation of China (Grant No. 16ZDA047 and 15CJL017), the Startup Foundation for Introducing Talent of NUIST (S8113097001), the Project Funded by the Flagship Major Development of Jiangsu Higher Education Institutions, the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the Top-notch Academic Programs Project of Jiangsu Higher Education Institutions.

References

[1] World Wind Energy Association. World wind energy report; 2017.
[2] Tasnim S, Rahman A, Oo AMT, Haque ME. Wind power prediction in new stations based on knowledge of existing stations: a cluster based multi-source domain adaptation approach. Knowl-Based Syst 2018;145:15–24.
[3] Tasnim S, Rahman A, Oo AMT, Haque ME. Autoencoder for wind power prediction. Renew: Wind Water Solar 2017;4(1):1–11.
[4] Zhao WG, Wei YM, Su ZY. One day ahead wind speed forecasting: a resampling-based approach. Appl Energy 2016;178:886–901.
[5] Yuan XH, Tan QX, Lei XH, Yuan YB, Wu XT. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy 2017;129:122–37.
[6] Wang C, Zhang HL, Fan WH, Fan XC. A new wind power prediction method based on chaotic theory and Bernstein Neural Network. Energy 2016;117:259–71.
[7] Hu JM, Wang JZ, Xiao LQ. A hybrid approach based on the Gaussian process with t-observation model for short-term wind speed forecasts. Renew Energy 2017;114:670–85.
[8] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Appl Energy 2011;88(4):1405–14.
[9] Lydia M, Suresh Kumar S, Immanuel Selvakumar A, Edwin Prem Kumar G. Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers Manage 2016;112:115–24.
[10] Cadenas E, Rivera W. Wind speed forecasting in the south coast of Oaxaca, Mexico. Renew Energy 2007;32(12):2116–28.
[11] Ren C, An N, Wang J, Li L, Hu B, Shang D. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. Knowl-Based Syst 2014;56:226–39.
[12] Zhang C, Wei H, Xie L, Shen Y, Zhang K. Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework. Neurocomputing 2016;205:53–63.
[13] Santamaría-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: a method based on support vector regression. Renew Energy 2016;85:790–809.
[14] Salcedo-Sanz S, Perez-Bellido AM, Ortiz-Garcia EG, Portilla-Figueras A, Prieto L, Paredes D. Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction. Renew Energy 2009;34:1451–7.
[15] Song JJ, Wang JZ, Lu HY. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. Appl Energy 2018;215:643–58.
[16] Wang JZ, Du P, Niu T, Yang WD. A novel hybrid system based on a new proposed algorithm-Multi-Objective Whale Optimization Algorithm for wind speed forecasting. Appl Energy 2017;208:344–60.
[17] Zhao J, Guo YL, Xiao X, Wang JZ, Chi DZ, Guo ZH. Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method. Appl Energy 2017;197:183–202.
[18] Salcedo-Sanz S, Perez-Bellido AM, Ortiz-Garcia EG, Portilla-Figueras A, Prieto L, Correoso F. Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks. Neurocomputing 2009;72:1336–41.
[19] Salcedo-Sanz S, Pastor-Sanchez A, Prieto L, Blanco-Aguilera A, Garcia-Herrera R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization – extreme learning machine approach. Energy ConversManage 2014;87:10–8.
[20] Salcedo-Sanz S, Cornejo-Bueno L, Prieto L, Paredes D, Garcia-Herrera R. Feature selection in machine learning prediction systems for renewable energy applications. Renew Sustain Energy Rev 2018;90:728–41.
[21] Kiplangat DC, Asokan K, Kumar KS. Improved week-ahead predictions of wind speed using simple linear models with wavelet decomposition. Renew Energy 2016;93:38–44.
[22] Zhang C, Wei H, Zhao J, Liu T, Zhu T, Zhang K. Short-term wind speed forecasting using empirical mode decomposition and feature selection. Renew Energy 2016;96:727–37.
[23] Liu H, Tian HQ, Li YF. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, Wavelet Packet-MLP and Wavelet Packet-ANFIS for wind speed predictions. Energy Convers Manage 2015;89:1–11.
[24] Wang HZ, Wang GB, Li GQ, Peng JC, Liu YT. Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl Energy 2016;182:80–93.
[25] Kuremoto T, Kimura S, Kobayashi K, Obayashi M. Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing 2014;137:47–56.
[26] Wang H, Li G, Wang G, Peng J, Jiang H, Liu Y. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[27] Wang JJ, Zhang WY, Wang JZ, Han TT, Kong LB. A novel hybrid approach for wind speed prediction. Inf Sci 2014;273:304–18.
[28] Wang Y, Wang J, Wei X. A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: a case study of wind farms in northwest China. Energy 2015;91:556–72.
[29] Shi J, Guo JM, Zheng SY. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renew Sustain Energy Rev 2012;16:3471–80.
[30] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62:531–44.
[31] Wang JJ, Wang YF, Li YN. A novel hybrid strategy using three-phase feature extraction and a weighted regularized extreme learning machine for multi-step ahead wind speed prediction. Energies 2018;11(2):321.
[32] Zhang F, Liu Y, Chen CJ, Li YF, Huang HZ. Fault diagnosis of rotating machinery based on kernel density estimation and Kullback-Leibler divergence. J Mech Sci Technol 2014;28(11):4441–54.
[33] Yan RQ, Gao RX. Rotary machine health diagnosis based on empirical mode decomposition. J Vib Acoust 2008;130(2):21007.
[34] Richman Joshua S, Randall Moorman J. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278:2039–49.
[35] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W, Woo W. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Adv Neural Inform Process Syst 2015;28:802–10.
[36] Liu HP, Erdem E, Shi J. Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. Appl Energy 2011;88(3):724–32.
[37] Bollerslev T, Wooldridge JM. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econ Rev 1992;11(2):143–72.