

Received July 20, 2020, accepted August 5, 2020, date of publication August 10, 2020, date of current version August 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015336

Short-Term Wind-Speed Forecasting Based on Multiscale Mathematical Morphological Decomposition, K-Means Clustering, and Stacked Denoising Autoencoders

WEICHAO DONG^{1,2}, HEXU SUN^{1,2}, (Senior Member, IEEE), ZHENG LI², (Member, IEEE), JINGXUAN ZHANG^{1,3}, (Graduate Student Member, IEEE), AND HUIFANG YANG^{1,4}

¹School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

²School of Electrical Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

³School of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, China

⁴Academic Affairs Office, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

Corresponding authors: Hexu Sun (hxsun@hebut.edu.cn) and Zheng Li (lizheng@hebut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51877070 and Grant 51577048, in part by the Natural Science Foundation of Hebei Province under Grant E2018208155 and Grant E2018210044, in part by the Talent Engineering Training Support Project of Hebei Province under Grant A201905008, in part by the National Engineering Laboratory of Energy-Saving Motor and Control Technique, Anhui University, under Grant KFKT201901, in part by the Key Project of Scientific and Technological Research in Universities of Hebei Province under Grant ZD2018228, and in part by the Hebei Education Department Fund under Grant QN2020433.

ABSTRACT Wind energy plays an increasingly important role in economic development. In this study, we propose a hybrid short-term wind-speed forecasting model comprising multiscale mathematical morphological decomposition (MMMD), K-means clustering algorithm, and stacked denoising autoencoder (SDAE) networks. First, in contrast to traditional signal-decomposing tools, the original wind-speed sequence is decomposed into a series of subsequences with different frequencies and fluctuant levels using the adaptive multiscale mathematical morphological algorithm directly in the time domain. The signal does not need to be transferred from the time domain to the frequency domain; hence, the accuracy can be considerably improved. Moreover, this is the first study that uses a time domain signal-decomposing tool in a hybrid wind forecasting model. Next, the data are split into different clusters of similar frequencies and fluctuant level subsequences using the K-means algorithm. The characteristics of each cluster are then captured using the SDAE as the core forecasting unit. Finally, the predictions of all subsequences are aggregated to obtain the final wind speed. The data from two real wind turbines are used to evaluate the performance of the proposed model, and the forecasting results are compared with five different benchmark models, namely, backpropagation neural network (BPNN), stacked denoising autoencoder (SDAE), mathematical morphology-backpropagation, mathematical morphology-SDAE, and K-means-SDAE for multiple scales, and two novel hybrid wind forecasting models namely, wavelet transform (WT)-K-means-SDAE and variation mode decomposition (VMD)-K-means-long short-term memory networks (LSTMs). The results of the comparison demonstrate that the proposed model provides a short-term wind-speed forecasting method whose prediction accuracy decreases with time; however, the proposed model achieves a better performance in comparison with other exiting models. At same time, the proposed model significantly increases the prediction accuracy of wind-speed forecasting and can be a reference for future research in this area.

INDEX TERMS K-means clustering, multiscale mathematical morphological decomposition, short-term wind-speed forecasting, stacked denoising autoencoders.

I. INTRODUCTION

Developing a low-carbon economy and striving for sustainable development has become a common aspiration of a

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

rapidly progressing human society. Recent developments in renewable energy have made wind the third-largest power source after coal and hydropower. However, wind power is naturally characterized by rapid fluctuations and intermittency that severely restrict its large-scale development. Therefore, formulating a reasonable wind power scheduling

plan based on accurate prediction and making it cooperate with traditional energy supplies are the most important steps to realize wind power scaling and regularization [1]–[3].

Wind-speed forecasting forms the basis of wind power forecasting and has been studied separately. The wind speed not only characterizes abundant noise, nonlinearity, and non-stationarity, but also, more importantly, presents wideband multiscale characteristics. Therefore, extracting robust and stable characteristics from complex wind-speed signals is quite challenging. The current level of wind-speed prediction is insufficient to meet the requirements of actual engineering applications [4]. Many recent studies focused on wind-speed forecasting models at multiple scales [5]. These forecasting models can be separated into very short-term, short-term, medium-term, and long-term forecasting based on time scales. Very short-term forecasting ranges from a few seconds to 30 min ahead. Short-term forecasting ranges from 30 min to 6 h ahead. Medium-term forecasting ranges from 6 h to 1 day ahead. Lastly, long-term forecasting ranges from 1 day to 1 week or more ahead. These models can be further classified into physical, statistical, intelligent, and hybrid models based on forecasting approaches [6], [7].

Wind-speed prediction is performed by depicting detailed physical discretion of the atmosphere using physical approaches. The most common method is numerical weather prediction (NWP), which has been widely adopted to predict wind-speed conditions. NWP predicts the atmospheric motion state and weather phenomena for a certain period in the future by solving the equations of fluid mechanics and thermodynamics describing the weather evolution process through numerical calculation. However, the computation process is time-consuming, and the performance suffers for short-term predictions [8], [9].

Statistical methods use statistical theory to establish the forecasting model based on historical data. Such methods are easy to model and are not based on any predefined conditions. The conventional model is an autoregressive integrated moving average (ARIMA) model. Yatiyana *et al.* [10] constructed a wind power forecasting model based on ARIMA time series. Their method achieved good prediction results, but the prediction process was time-consuming owing to a single ARIMA model being employed to simultaneously complete multiple tasks. In [11], many different statistical models are proposed to predict wind speed and power. However, statistical methods are not suitable to deal with nonlinear patterns.

Intelligent models have demonstrated good performance in processing nonlinear patterns, and artificial neural networks (ANNs) are the most prevalent predictors in intelligent models. Peng *et al.* [12] proposed using a stacked denoising autoencoder (SDAE) model to predict wind power. They used the data from real wind turbines to evaluate the performance of the proposed model compared to two different models, namely backpropagation neural network (BPNN) and support vector machines, in terms of prediction results. The comparison illustrated that SDAE has a more robust prediction ability to deal with nonlinear data. A bidirectional gated recurrent

unit-based deep learning model demonstrated superior wind power forecasting in [13], and the results were verified using real data from a wind farm. Wang *et al.* [14] introduced a deep belief network (DBN) on a multi-dimensional phase space to predict wind power. Compared with other benchmark models, using data from a real wind turbine, the DBN model showed better ability when processing nonlinear systems. Lin *et al.* [15] used an improved DBN model with genetic algorithms for wind-speed prediction with increased accuracy.

Conventional ANNs have serious drawbacks such as falling into local minima and overfitting. Several hybrid models combining different approaches have been widely applied to solve these problems. The signal-decomposing technique has been widely used to construct hybrid models. Sun *et al.* [16] introduced a hybrid model integrating variation mode decomposition (VMD), K-means clustering, and long short-term memory networks to forecast wind power. The prediction result demonstrated its superior performance when compared with six different benchmark models. In [17]–[19], wavelet transform (WT) was used as a signal-decomposing tool for the input signal. However, both VMD and WT accomplish data processing in the frequency domain, which requires transferring the data from the time domain to the frequency domain and back, thereby increasing signal error. Chen *et al.* [20] proposed wind-speed prediction model based on multiscale mathematical morphology and support vector regression. The mathematical morphology algorithm can decompose wind-speed sequences into a series of subsequences directly in the time domain with large improvements in accuracy. However, the structure of their entire model is simple and offers poor prediction performance.

This study introduces a novel short-term wind-speed forecasting model based on multiscale mathematical morphological decomposition, K-means clustering, and SDAE. The contributions of this study can be summarized as follows:

1. A multiscale mathematical morphology algorithm is used to decompose wind-speed sequences into a series of subsequences in the time domain.
2. A K-means clustering algorithm is applied to classify the data into different clusters with similar frequencies and fluctuant level subsequences.
3. SDAE, a deep learning method, is proposed as the core forecasting unit to capture each cluster's characteristics.
4. The predictions of all subsequences are aggregated to obtain the final predicted wind speed.
5. Data from two real wind turbines are used to comparatively evaluate the performances of the proposed model, five different benchmark models, and two novel hybrid wind forecasting models.

The remainder of this article is organized as follows. Section II presents the methodologies used in this study. Section III evaluates the case studies of the proposed wind-speed forecasting framework. Finally, Section IV presents concluding remarks.

II. METHODOLOGIES

A. MULTISCALE MATHEMATICAL MORPHOLOGICAL DECOMPOSITION

Mathematical morphology (MM) is a nonlinear analysis method based on strict mathematical theory. The basic principle is to use a probe, called a structuring element (SE), to move through the signal and perform basic operations to extract useful feature information. The basic operations of the MM include dilation, erosion, opening, and closing [21]–[23].

In the decomposition process, the SE has a function similar to that of a filtering window for general signal processing. The signal is extracted when the shape of the signal matches the shape of the SE. In practical applications, the shape of the SE can be determined according to the target signal characteristics. The common SE types include linear, triangular, circular, and cosine.

Let $f(n)$ be the pending signal, which is the discrete function over the domain $D_f = \{0, 1, 2, \dots, N\}$ and let $g(n)$ be the 1-D SE, which is the discrete function over the domain $D_g = \{0, 1, 2, \dots, P\}$. Both N and P are integers, and $N \geq P$. The morphological operators (i.e., erosion and dilation) can be defined as

$$(f \oplus g)(n) = \max \{f(n - m) + g(m), n \in D_f, m \in D_g\} \quad (1)$$

$$(f \ominus g)(n) = \min \{f(n + m) - g(m), n \in D_f, m \in D_g\} \quad (2)$$

where \oplus and \ominus denote the erosion and dilation operators, respectively. Two other basic morphological operators, namely, the opening and closing, can be further defined based on dilation and erosion, as follows:

$$(f \circ g)(n) = [(f \ominus g) \oplus g](n) \quad (3)$$

$$(f \bullet g)(n) = [(f \oplus g) \ominus g](n) \quad (4)$$

where \circ is the opening operator, and \bullet is the closing operator.

The nonlinear and nonstationary characteristics of the wind determining the wind speed have multi-time scales. We introduce a multiscale morphology analysis based on the traditional single-scale morphology filter. By defining different SE sizes, we perform omnidirectional scans of the wind-speed curve and extracts fluctuation characteristics at different scales. In this case, we can depict the morphological characteristics of the wind-speed curve hierarchically.

Let f , g , and T denote a discrete pending signal, the SE of the MM, and multiple morphological operator, respectively. The multiscale morphological operation is based on the set $\{T_s | s > 0, s \in \mathbb{Z}\}$, where s is a positive integer representing the scale of the SE, and

$$T_s(f) = sT\left(\frac{f}{s}\right), \quad s > 0 \quad (5)$$

Similarly, multiscale MM operator erosion and dilation can be respectively expressed as follows:

$$(f \ominus g)_s(n) = s \left[\left(\frac{f}{s} \right) \ominus g \right](n) = f \ominus sg(n) \quad (6)$$

$$(f \oplus g)_s(n) = s \left[\left(\frac{f}{s} \right) \oplus g \right](n) = f \oplus sg(n) \quad (7)$$

where $sg = g \oplus g \oplus \dots \oplus g$ ($s-1$ times). The open and closed operations, respectively, of the multiscale morphology, are defined as

$$(f \circ g)_s(n) = f \ominus g \ominus g \ominus \dots \ominus g \oplus g \oplus g \oplus \dots \oplus g \quad (8)$$

$$(f \bullet g)_s(n) = f \oplus \underbrace{g \oplus g \oplus \dots \oplus g}_s \ominus \underbrace{g \ominus g \ominus \dots \ominus g}_s \quad (9)$$

The signal decomposition process can be regarded as a multiple filtering process. The SEs of different scales can adapt to different signal shapes. Before the signal decomposition, two preparatory works need to be performed: choosing suitable size of the SE and filter.

The size of the SE is determined by the signal length and height. The local peak values of the original signal $X = \{x_n | n = 0, 1, 2, \dots, N - 1\}$ are calculated, where N is the signal length. Let $P = \{p_n | n = 0, 1, 2, \dots, N_p\}$ be the series of peaks, where N_p is the number of peaks. The peak interval I is defined as $I = \{i_n | i_n = p_{n+1} - p_n, n = 0, 1, 2, \dots, N_p - 1\}$. L_{\max} and L_{\min} are defined as the minimum and maximum lengths of the SEs, respectively:

$$L_{\max} = \lfloor (I_{\max} - 1) / 2 \rfloor \quad (10)$$

$$L_{\min} = \lceil (I_{\min} - 1) / 2 \rceil \quad (11)$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ represent the operators rounded toward infinity and rounded toward minus infinity, respectively.

The length and height of the SEs at different scales j are then obtained as follows:

$$L_j = L_{\min} + j, \quad (j = 0, 1, 2, \dots, L_{\max} - L_{\min}) \quad (12)$$

$$h_j = \delta * [h_{\min} + j * (h_{\max} - h_{\min}) / (L_{\max} - L_{\min})] \quad (13)$$

where δ ($0 \leq \delta \leq 1$) is the proportionality coefficient.

A suitable filter must be selected afterward. Alternating and hybrid alternating filters are the most common type of filters used in the multiscale morphology. Alternating filters are also known as morphologic open–close (OC) filters and morphologic closed–open (CO) filters. These filters are composed of morphological open and closed operations cascaded in different sequences and are written as follows:

$$h_{oc}(f)_s(n) = ((f \circ g)_s \bullet g) \quad (14)$$

$$h_{co}(f)_s(n) = ((f \bullet g)_s \circ g) \quad (15)$$

To compensate for the shortcomings of the morphological OC and CO filters, they are often combined to form a hybrid alternating filter known as the OCCO filter, expressed as

$$h_{occo} = 0.5 * [h_{oc}(f)_s(n) + h_{co}(f)_s(n)] \quad (16)$$

In this study, the signal characteristics of each scale can be effectively extracted. Moreover, more ideal morphological features can be obtained using the weighted multiscale morphology filter (WMMF).

Let the scale of SE be $S = \{S_1, S_2, \dots, S_k\}$. We can realize the OCCO morphological filtering $h_{occo}(f)_{s_i}$ at k different effective scales, where $i = 1, 2, \dots, k$. The WMMF is defined as,

$$h_{WMMF}(f)_s(n) = \sum_{i=1}^k \omega_{s_i} \times h_{occo}f_{s_i}(n) \quad (17)$$

$$\omega_{s_i} = \frac{\sigma_{s_i}^2}{\sum_{i=1}^k \sigma_{s_i}^2} \quad i = 1, 2, \dots, k \quad (18)$$

where ω_{s_i} and $\sigma_{s_i}^2$ represent the weight factors and variances of each SE, respectively. The corresponding ω_{s_i} is also small because of the weak denoising ability of small-scale SEs. The WMMF combines filters with different-scale SEs, ensuring the preservation of the original signal characteristics as much as possible. Fig. 1 shows the raw input wind-speed processing after completion of the two preparatory works.

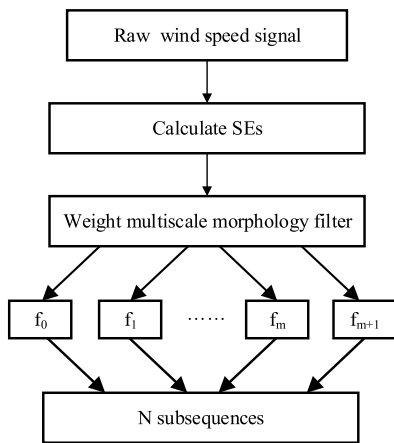


FIGURE 1. Raw wind speed processing.

Step 1: Choose triangular-type SEs to finish the filtering task, then design the SEs in different scales, $G_j = (SEtri)_j$. From (12) and (13), we obtain

$$(SEtri)_j = h_j * [01 \dots L_{j-1}L_jL_{j-1} \dots 10] \quad (19)$$

Step 2: Let the input signal be denoted as F ; the output $y_j(x)$ at scale j is obtained using (17).

Step 3: We obtain $j(j = 0, 1, 2, \dots, m)$ differently scaled outputs according to Step 2.

$$F = \sum_{i=1}^k f_k = \begin{cases} f_0 = F - y_0 \\ f_j = y_{j-i} - y_j, 1 \leq j \leq m \\ f_{m+1} = y_m \end{cases} \quad (20)$$

After decomposition, the original signal sequence is decomposed into a series of $m + 1$ detail components, $f_0 - f_m$ and a principal component f_{m+1} to obtain a total of $m + 2$ subsequence layers. According to (19) and (20), m depends on j ,

which is the scale coefficient of the SEs. Equations (10) – (12) show that j depends on the peak interval of the original input signal. In other words, the number of layers in subsequence depends on the original input signal.

B. K-MEANS ALGORITHM

K-means is a simple and classical clustering algorithm based on distance. The idea of K-means was first presented by Hugo Steinhaus in 1957 [24]. As a data-mining approach, the K-means algorithm automatically groups the input data into the corresponding predefined clusters by minimizing the distance function in an unsupervised manner. The points in the cluster are connected as closely as possible. The distance between the clusters is as large as possible [25], [26]. Figure 2 illustrates the procedure of the K-means algorithm [27].

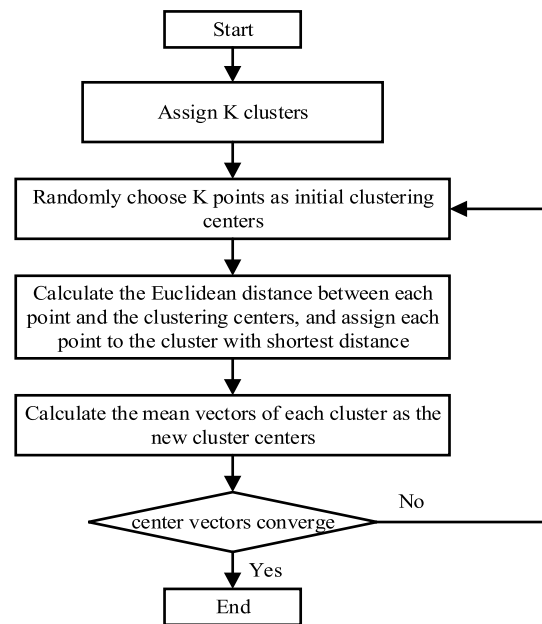


FIGURE 2. Flowchart of the K-means algorithm.

Step 1: Determine the k different clusters in advance.
 Step 2: Randomly choose K points as the initial clustering centers.
 Step 3: Calculate the Euclidean distance $D(x, y)$ between each point x_i and the clustering centers y_i , and assign each point to the cluster with the shortest distance using the Euclidean distance formula as

$$D(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (21)$$

Step 4: Redefine the cluster centers by calculating the mean vectors based on the following equation:

$$\mu_j = \frac{1}{N_k} \sum_{i=1}^{N_k} data_i^k \quad (22)$$

where μ_j is the center vector of the j^{th} cluster; $data_i^k$ is the i^{th} data in cluster k ; and N_k is the sample amount of each cluster.

Step 5: Repeat steps 3 and 4 until the center vectors converge.

The K-means algorithm can be used on each subsequence of the decomposition result F to cluster the wind-speed data into different categories.

C. STACKED DENOISING AUTOENCODER

The autoencoder (AE) is an ANN with three layers—the input, hidden, and output layers—which are mainly used for dimensionality reduction or feature extraction. The hidden layer leads to dimensionality reduction and can help reconstruct the input data. Fig. 3 displays the basic structure of an autoencoder [28],[29].

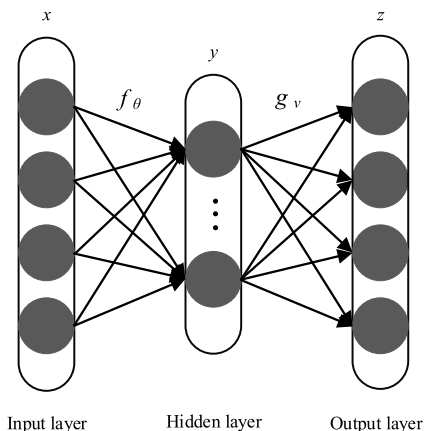


FIGURE 3. General autoencoder structure.

The autoencoder maps the input data into a hidden representation f using

$$y = f_{\theta}(x) = \sigma(Wx + b) \tag{23}$$

where σ is the sigmoidal activation function. $\theta_f = \{W, b\}$ is the parameter set containing the transformed weight matrix W , and bias vector b . $f_{\theta}(x)$ is a function that encodes the features from the input layer to the hidden layer. Then, the hidden layer maps the feature to the output layer based on

$$z = g_v(y) = \sigma(W'y + b') \tag{24}$$

where $v_g = \{W', b'\}$ is the parameter set containing the transformed weight matrix W' and bias vector b' . The function $g_v(y)$ decodes the feature back to the output layer. The decoding weight matrix $W' = W^T$. The AE system is trained by minimizing the loss function as follows:

$$L_{\theta,v} = \arg \min_{\theta,v} \frac{1}{n} \sum_{i=1}^n \|x^{(i)} - g_v(y)\|^2 \tag{25}$$

where n is the number of samples. However, the structure of AE is very simple, leading to overfitting and thereby reducing the efficiency of the feature extraction.

The denoising autoencoder (DAE) is an improved version of the AE, with similar network structures and operating

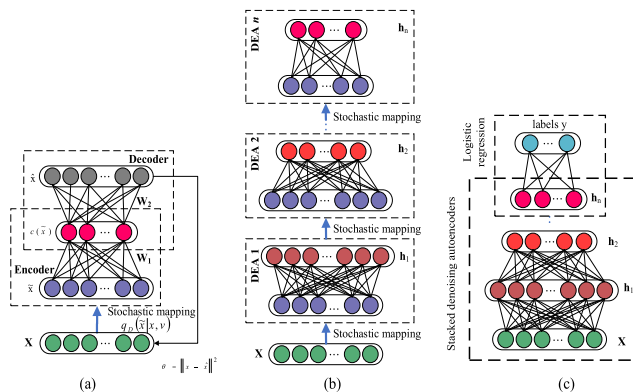


FIGURE 4. SDAE. (a) Structure of DAE. (b) Unsupervised pretraining of SDAE. (c) Supervised fine-tuning of SDAE.

process. However, DAE reconstructs the input signal by corrupting one to get a more robust system (Fig. 4(a)) [30].

The original input x is stochastically corrupted to \tilde{x} , and the encoding process of the corrupted input is

$$c(\tilde{x}) = \text{sig}(W'\tilde{x} + b') \tag{26}$$

An SDAE is made up of multiple DAEs, with the aim of building a deep architecture. In general, the SDAE has two learning steps: an unsupervised pretraining step and a supervised fine-tuning step (Figs. 4(b) and (c), respectively). The learning procedure starts with a greedy layer-wise pretraining procedure. Each DAE layer is trained in the same manner. The output of each DAE is the input of the next DAE. The first step performs unsupervised training of each DAE layer separately to minimize the error between the input and reconstruction results. After pretraining of the DAEs, all hidden layers are trained. A logistic regression layer is then added on top of the hidden layers; subsequently, the data with labels are used to fine-tune the network further through a BPNN algorithm.

D. OPERATING PROCEDURE OF PROPOSED MODEL

Fig. 5 demonstrates the framework of the MMD–K-means–SDAE model. The specific details are presented below.

(1) For the given training dataset, the input signals are decomposed into a series of subsequences with different frequencies and fluctuant levels using the MMD. The triangular SE shape and WMMF filters are then selected.

(2) The K-means algorithm is used for each subsequence of the decomposition result F to cluster the wind-speed data into several categories. The Euclidean distances between each point and clustering centers are calculated, and each point is assigned to the cluster with the shortest distance.

(3) Optimal SDAE models are established based on the clustering result. The predictions of all subsequences from the hybrid model are then aggregated to obtain the final result.

The model proposed in this study is for short-term wind-speed forecasting; therefore, the effective prediction time scale is 30 min to 2 h. Beyond this prediction range, the prediction sensitivity decreases with time. When the

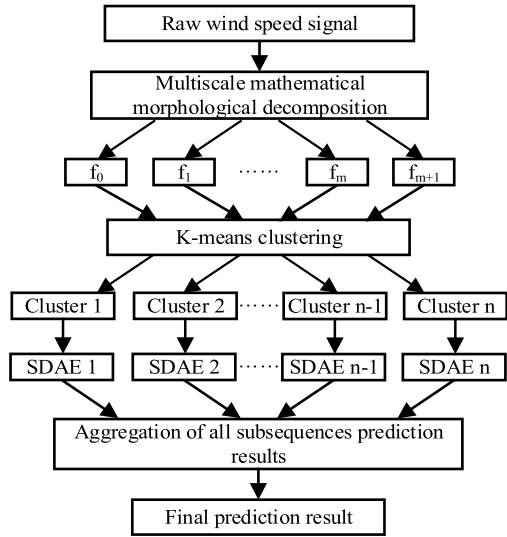


FIGURE 5. Operating procedure of the proposed MMMD-K-means-SDAE model.

time scale exceeds 48 h, the model will not provide a valid prediction.

III. CASE STUDY

The experiments were run on a platform with the following configuration: AMD Ryzen 2600 Six-Core Processor, 3.40 GHz, 16.0 GB RAM. The models were applied on Python 3.7, TensorFlow-GPU 1.15.0, and Keras 2.1.4. A comparison of the total computation time of different models is shown in Table 1. As a three-layer hybrid model, the computation time of the proposed model is 513 s; therefore, the computation efficiency of the proposed method is very high compared to other models.

TABLE 1. Comparison of total computation time.

Model name	Time (s)
MMMD-K-means-SDAE	513
VMD-K-means-LSTM	542
WT-K-means-SDAE	556
MMMD-SDAE	402
K-means-SDAE	361
MMMD-BPNN	300
SDAE	206
BPNN	169

A. DATA DESCRIPTION

Data from two randomly selected wind turbines collected for the last quarter of 2017 from Hebei Province, China were used. Approximately 25,000 SCADA data were available for each wind turbine. Of these, 15,000 data units were randomly selected as the training data, and the remainder were used as the testing data. The proposed method can be universally

applied to all wind farm scenarios; therefore, there are no special requirements for the datasets. The cut-in wind speed of the wind turbine was 2.5–3 m/s, the cut-off wind speed was 25 m/s, and the rated speed of the wind turbine was set to 11 m/s. The mean relative error (MRE), root-mean-square error (RMSE), and mean absolute percentage error (MAPE) were selected as indicators to compare the proposed model with the benchmark models. The MRE represents the average of the absolute error between the predicted and actual values. It is a linear function, and all individual differences have an equal weight on average. The RMSE indicates the sample standard deviation between the predicted value and the actual value, demonstrating the dispersion degree of the predicted value. The MAPE is similar to the MRE, which represents the percentage of the error in the actual value. All three indicators are used to evaluate prediction accuracy, which decreases with an increase in the values of these three indicators. They can be calculated as follows:

$$MRE = \frac{\sum_{i=1}^N (|x(t) - \hat{x}(t)| / x_r)}{N} \tag{27}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N [x(t) - \hat{x}(t)]^2}{N}} \tag{28}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{x(t) - \hat{x}(t)}{x(t)} \right| \tag{29}$$

where $x(t)$ is the actual data; x_r is the rated value; $\hat{x}(t)$ is the predicted data; and N is the number of forecasting samples.

For further comparison of the performances of two different models, the improved performance of the MRE (P_{MRE}), RMSE (P_{RMSE}), and MAPE (P_{MAPE}) were introduced and calculated as

$$P_{MRE} = \frac{|MRE_1 - MRE_2|}{MRE_1} \tag{30}$$

$$P_{RMSE} = \frac{|RMSE_1 - RMSE_2|}{RMSE_1} \tag{31}$$

$$P_{MAPE} = \frac{|MAPE_1 - MAPE_2|}{MAPE_1} \tag{32}$$

The P_{MRE} , P_{RMSE} , and P_{MAPE} represent the result of comparing two different forecasting models. The P_{MRE} , P_{RMSE} , and P_{MAPE} are presented as percentages; thus, the performance of these two models can be further analyzed.

B. DECOMPOSITION RESULTS OF THE ORIGINAL WIND-SPEED SIGNAL

The first step in the prediction process is the decomposition of the original wind-speed sequences into several subsequences. The wind-speed signal was decomposed into seven and five subsequences for wind turbines #1 and #2, respectively, based on the original input signal. The testing data of the wind-speed sequences were obtained from the two wind turbines (Fig. 6). Figs. 7(a) and (b) illustrate the decomposition results of the two wind turbines using the MMMD technique. For wind turbine #1, f_0 – f_5 are the detail components,

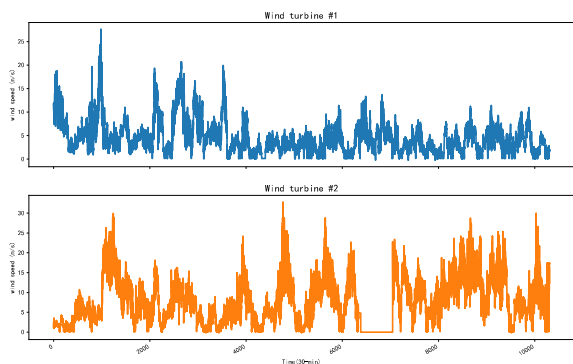


FIGURE 6. Testing data of thirty-minute wind speed sequences.

and f_6 is the principal component. For wind turbine #2, f_0 – f_3 are the detail components, and f_4 is the principal component.

C. EFFECT OF ESTIMATION OF THE PROPOSED MODEL

After signal decomposition, the wind signal was used to make 30 min, 1 h, and 2 h forecasting in the proposed model. The forecasting results were compared with those obtained using seven other approaches, namely, BPNN, SDAE, MMMD–BPNN, MMMD–SDAE, K-means–SDAE, WT-K-means–SDAE, and VMD-K-means-LSTM. Figs. 8–13 depict the comparison results with different time scales. Tables 2 and 3 present the estimated errors of the different models.

Figs. 8–13 show that the results from all forecasting models share the same characteristics. To further prove the performance of the proposed hybrid model, P_{MRE} , P_{RMSE} , and P_{MAPE} were introduced to compare the eight different models. Tables 4–13 present the comparison results for wind turbine #1, whereas Tables 14–23 show the comparison results for wind turbine #2. Using the proposed model as a reference, a comparison of the P_{MRE} , P_{RMSE} , and P_{MAPE} of all the other models for the two wind turbines is shown in Tables 24 and 25.

The following inferences can be drawn from the comparison results given in Tables 2–25.

1) SDAE VS. BPNN

The SDAE approach can provide more accurate forecasting results than the BPNN, as indicated by the P_{MRE} , P_{RMSE} , and P_{MAPE} for the 30 min, 1 h, and 2 h time scales for wind turbine #1. The improvements in P_{MRE} were 14.97, 22.75, and 19.20, respectively, whereas those in P_{RMSE} were 16.17, 21.47, and 20.72, respectively, and in P_{MAPE} were 7.79, 12.10, and 10.20, respectively. For wind turbine #2 at the same time scales, the SDAE model could improve P_{MRE} by 17.92, 14.81, and 14.56, respectively, P_{RMSE} by 17.17, 15.48, and 15.66, respectively, and P_{MAPE} by 9.40, 7.70, and 7.56 respectively. In other words, a deep learning network has a better performance than a traditional ANN due to avoidance of the local minimum and overfitting problems.

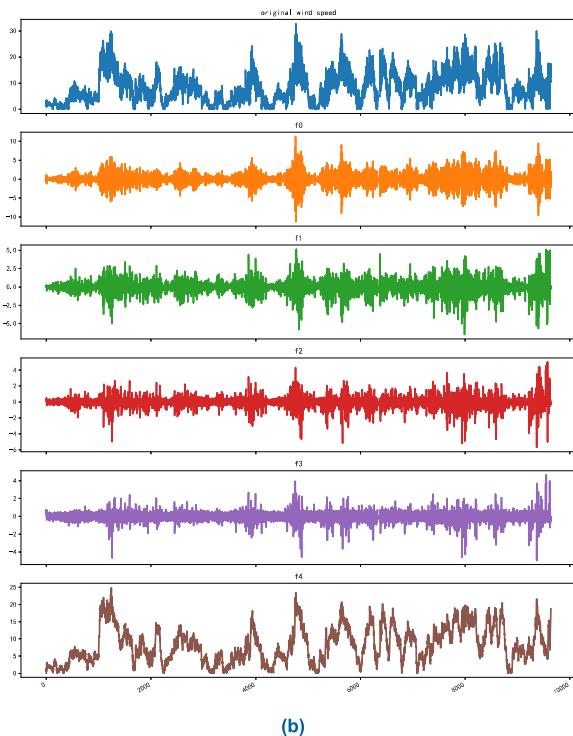
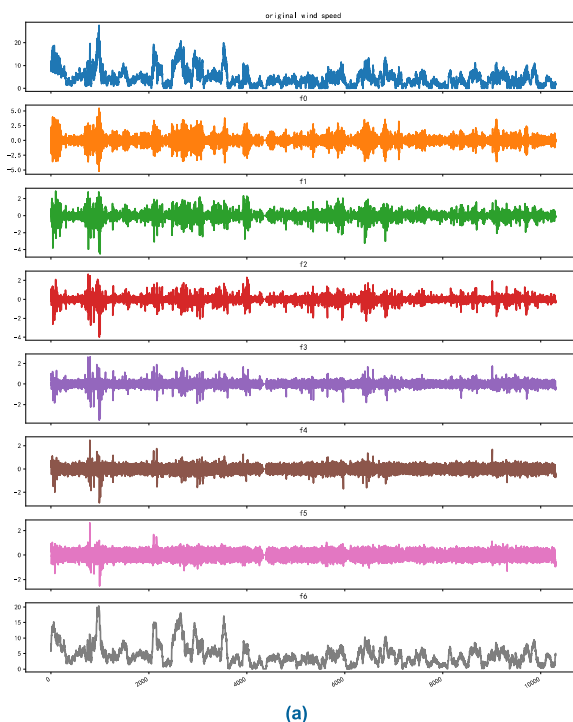


FIGURE 7. Thirty-minute decomposition results of the two wind turbines using the MMMD technique. (a) Thirty-minute decomposition results of wind turbine #1. (b) Thirty-minute decomposition results of wind turbine #2.

2) MMMD–SDAE VS. SDAE

The SDAE model demonstrates a stronger generalized ability when performing the MMM decomposition technique before forecasting. Wind turbine #1 had time scales of 30 min,

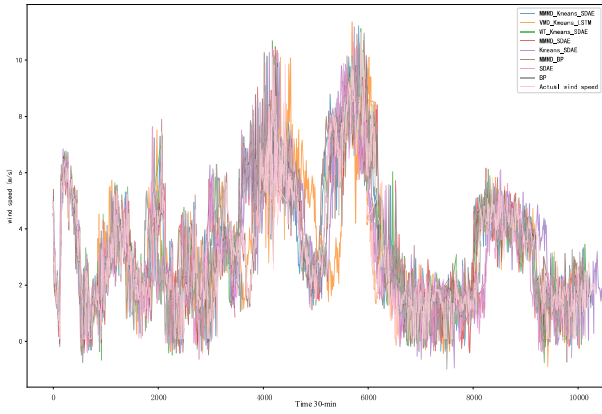


FIGURE 8. Comparison results of wind turbine #1 for the thirty-minute time scale.

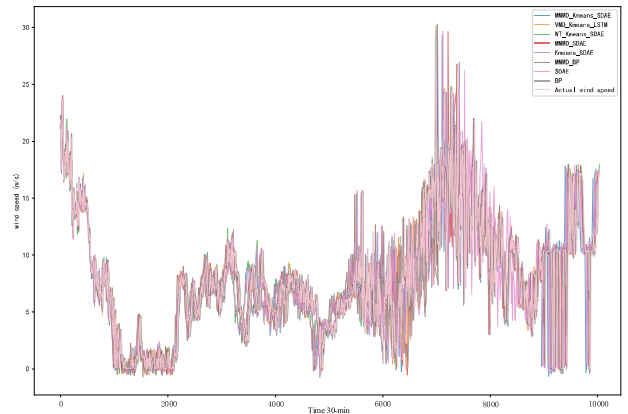


FIGURE 11. Comparison results of wind turbine #2 for the thirty-minute time scale.

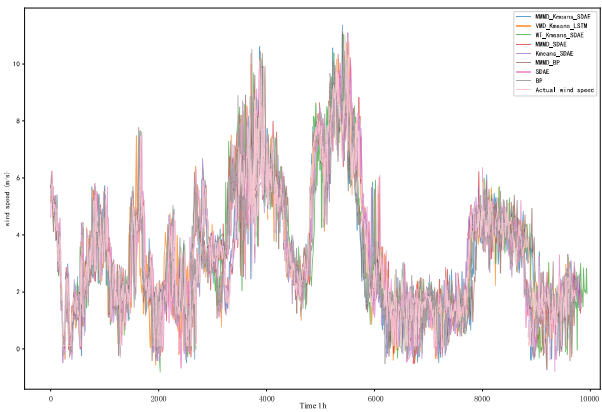


FIGURE 9. Comparison results of wind turbine #1 for the one-hour time scale.

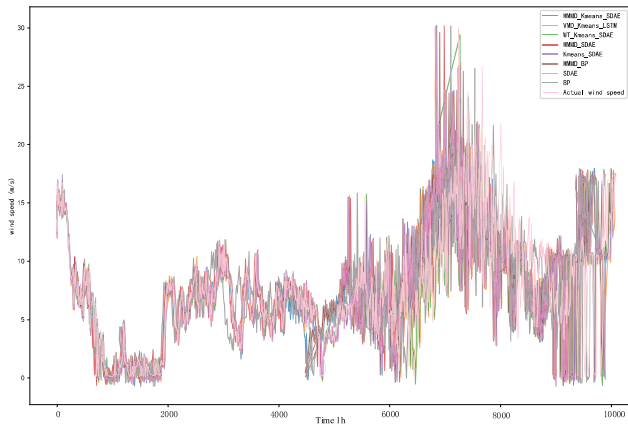


FIGURE 12. Comparison results of wind turbine #2 for the one-hour time scale.

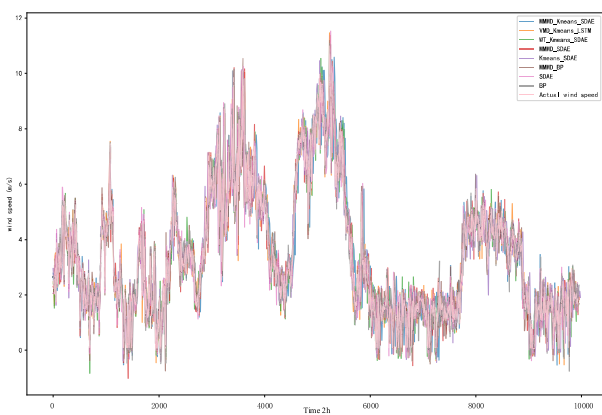


FIGURE 10. Comparison results of wind turbine #1 for the two-hour time scale.

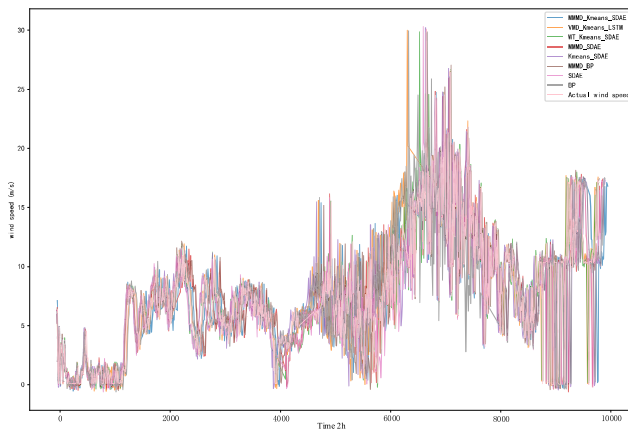


FIGURE 13. Comparison results of wind turbine #2 for the two-hour time scale.

1 h, and 2 h. The improvements in P_{MRE} were 72.01, 64.20, and 56.62, respectively, whereas those in P_{RMSE} were 70.84, 64.34, and 55.61, respectively, and those in P_{MAPE} were 47.16, 40.17, and 34.14, respectively. For wind turbine #2

at the same time scales, the MMMD–SDAE model could improve P_{MRE} by 52.75, 59.38, and 48.39, respectively, P_{RMSE} by 49.59, 57.48, and 48.32, respectively, and P_{MAPE} by 31.26, 36.27, and 28.16, respectively. The MMMD algorithm decomposes wind-speed sequences into a series of

TABLE 2. Comparison of forecasting errors between the proposed and benchmark models for wind turbine #1.

Forecasting approaches	Time scale	MRE	RMSE	MAPE (%)
BPNN approach	30 min	0.4342	5.983	16.103
	1 h	0.6501	8.842	19.704
	2 h	0.7076	9.823	20.557
SDAE approach	30 min	0.3692	4.983	14.849
	1 h	0.5022	6.944	17.318
	2 h	0.5717	7.788	18.478
MMMD–BPNN approach	30 min	0.2717	3.822	12.738
	1 h	0.4137	5.778	15.718
	2 h	0.4447	6.154	16.296
MMMD–SDAE approach	30 min	0.1031	1.453	7.846
	1 h	0.1798	2.476	10.362
	2 h	0.2480	3.457	12.170
K-means–SDAE approach	30 min	0.1899	2.643	10.649
	1 h	0.2882	3.980	13.119
	2 h	0.3383	4.644	14.213
MMMD–K-means–SDAE approach	30 min	0.0598	0.885	5.976
	1 h	0.1284	1.785	8.757
	2 h	0.1643	2.244	9.906
WT-K-means-SDAE approach	30 min	0.0697	1.009	6.452
	1 h	0.1486	1.986	9.420
	2 h	0.2158	2.651	11.352
VMD-K-means-LSTM approach	30 min	0.0691	1.005	6.423
	1 h	0.1475	1.975	9.385
	2 h	0.2135	2.648	11.292

subsequences directly in the time domain; hence, it can avoid errors during the signal transfer process and greatly increases the forecasting accuracy and stability.

3) K-MEANS–SDAE VS. SDAE

Relative to the traditional SDAE model, the accuracy of the prediction result will improve when adding the K-means clustering algorithm. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h. The improvements in P_{MRE} were 48.56, 42.61, and 40.83, respectively, whereas those in P_{RMSE} were 46.96, 42.68, and 40.37, respectively, and those in P_{MAPE} were 28.28, 24.25, and 23.08, respectively. For wind turbine #2 at the same time scales, the K-means–SDAE model could improve P_{MRE} by 35.39, 41.25, and 35.17, respectively, P_{RMSE} by 34.61, 39.69, and 33.45, respectively, and P_{MAPE} by 19.62, 23.35, and 19.48, respectively.

4) MMMD–K-MEANS–SDAE VS. SDAE

Compared to the SDAE approach, the proposed hybrid MMMD–K-means–SDAE produced a better forecasting performance. Wind turbine #1 had time scales of 30 min, 1 h, and 2 h. The improvements in P_{MRE} were 83.80, 74.43, and 71.26,

TABLE 3. Comparison of forecasting errors between the proposed and benchmark models for wind turbine #2.

Forecasting approaches	Time scale	MRE	RMSE	MAPE (%)
BPNN approach	30 min	0.4654	6.408	16.671
	1 h	0.6590	9.071	19.838
	2 h	0.6915	9.659	20.321
SDAE approach	30 min	0.3820	5.308	15.103
	1 h	0.5614	7.667	18.310
	2 h	0.5908	8.146	18.783
MMMD–BPNN approach	30 min	0.3054	4.210	13.504
	1 h	0.4288	5.980	16.002
	2 h	0.4915	6.934	17.132
MMMD–SDAE approach	30 min	0.1805	2.676	10.382
	1 h	0.2280	3.260	11.668
	2 h	0.3049	4.271	13.493
K-means–SDAE approach	30 min	0.2468	3.471	12.140
	1 h	0.3298	4.624	14.035
	2 h	0.3830	5.421	15.123
MMMD–K-means–SDAE approach	30 min	0.1451	2.322	9.308
	1 h	0.1686	2.553	10.034
	2 h	0.2183	3.163	11.417
WT-K-means-SDAE approach	30 min	0.1493	2.375	9.442
	1 h	0.2061	3.024	11.094
	2 h	0.2937	4.167	13.243
VMD-K-means-LSTM approach	30 min	0.1489	2.353	9.429
	1 h	0.2039	2.969	11.034
	2 h	0.2935	4.128	13.239

TABLE 4. Comparison between SDAE and BPNN for wind turbine #1.

Indices	SDAE vs. BPNN		
	30 min	1 h	2 h
P_{MRE} (%)	14.97	22.75	19.20
P_{RMSE} (%)	16.17	21.47	20.72
P_{MAPE} (%)	7.79	12.10	10.20

TABLE 5. Comparison between MMMD–SDAE and SDAE for wind turbine #1.

Indices	MMMD–SDAE vs. SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	72.01	64.20	56.62
P_{RMSE} (%)	70.84	64.34	55.61
P_{MAPE} (%)	47.16	40.17	34.14

respectively, whereas those in P_{RMSE} were 82.24, 74.29, and 71.19, respectively, and those in P_{MAPE} were 59.75, 49.43, and 46.39, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model can improve P_{MRE} by

TABLE 6. Comparison between K-means-SDAE and SDAE for wind turbine #1.

Indices \ Scales	K-means-SDAE vs. SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	48.56	42.61	40.83
P_{RMSE} (%)	46.96	42.68	40.37
P_{MAPE} (%)	28.28	24.25	23.08

TABLE 7. Comparison between MMD-K-means-SDAE and SDAE for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	83.80	74.43	71.26
P_{RMSE} (%)	82.24	74.29	71.19
P_{MAPE} (%)	59.75	49.43	46.39

TABLE 8. Comparison between MMD-K-means-SDAE and BPNN for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. BPNN		
	30 min	1 h	2 h
P_{MRE} (%)	86.22	80.25	76.78
P_{RMSE} (%)	85.20	79.81	77.16
P_{MAPE} (%)	62.89	55.56	51.81

TABLE 9. Comparison between MMD-K-means-SDAE and MMMD-BPNN for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. MMMD-BPNN		
	30 min	1 h	2 h
P_{MRE} (%)	77.99	68.96	63.05
P_{RMSE} (%)	76.84	69.11	63.54
P_{MAPE} (%)	53.08	44.28	39.21

TABLE 10. Comparison between MMMD-K-means-SDAE and MMMD-SDAE for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. MMMD-SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	42.00	28.59	33.75
P_{RMSE} (%)	39.09	27.91	35.09
P_{MAPE} (%)	23.84	15.49	18.60

TABLE 11. Comparison between MMMD-K-means-SDAE and K-means-SDAE for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. K-means-SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	68.51	55.45	51.43
P_{RMSE} (%)	66.52	55.15	51.68
P_{MAPE} (%)	43.84	33.25	30.31

62.02, 69.97, and 63.05, respectively, P_{RMSE} by 56.25, 66.70, and 61.17, respectively, and P_{MAPE} by 38.36, 45.19, and 39.21, respectively. The prediction results from the proposed

TABLE 12. Comparison between MMMD-K-Means-SDAE and WT-K-Means-SDAE for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. WT-K-means-SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	14.20	13.60	23.86
P_{RMSE} (%)	12.29	10.12	15.35
P_{MAPE} (%)	7.38	7.04	12.74

TABLE 13. Comparison between MMMD-K-means-SDAE and VMD-K-means-LSTM for wind turbine #1.

Indices \ Scales	MMMD-K-means-SDAE vs. VMD-K-means-LSTM		
	30 min	1 h	2 h
P_{MRE} (%)	13.45	12.95	23.04
P_{RMSE} (%)	11.94	9.62	15.26
P_{MAPE} (%)	6.96	6.69	12.27

TABLE 14. Comparison between SDAE and BPNN for wind turbine #2.

Indices \ Scales	SDAE vs. BPNN		
	30 min	1 h	2 h
P_{MRE} (%)	17.92	14.81	14.56
P_{RMSE} (%)	17.17	15.48	15.66
P_{MAPE} (%)	9.40	7.70	7.56

TABLE 15. Comparison between MMMD-SDAE and SDAE for wind turbine #2.

Indices \ Scales	MMMD-SDAE vs. SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	52.75	59.38	48.39
P_{RMSE} (%)	49.59	57.48	48.32
P_{MAPE} (%)	31.26	36.27	28.16

TABLE 16. Comparison between K-means-SDAE and SDAE for wind turbine #2.

Indices \ Scales	K-means-SDAE vs. SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	35.39	41.25	35.17
P_{RMSE} (%)	34.61	39.69	33.45
P_{MAPE} (%)	19.62	23.35	19.48

MMMD-K-means-SDAE model were closer to the actual values.

5) MMMD-K-MEANS-SDAE VS. BPNN

Compared to the BPNN approach, the proposed hybrid MMMD-K-means-SDAE produces a better forecasting performance. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h. The improvements in P_{MRE} were 86.22, 80.25, and 76.78, respectively, whereas those in P_{RMSE} were 85.20, 79.81, and 77.16, respectively, and those in P_{MAPE}

TABLE 17. Comparison between MMMD–K-means–SDAE and SDAE for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	62.02	69.97	63.05
P_{RMSE} (%)	56.25	66.70	61.17
P_{MAPE} (%)	38.36	45.19	39.21

TABLE 18. Comparison between MMMD–K-means–SDAE and BPNN for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. BPNN		
	30 min	1 h	2 h
P_{MRE} (%)	68.82	74.42	68.43
P_{RMSE} (%)	63.76	71.86	67.25
P_{MAPE} (%)	44.16	49.41	43.81

TABLE 19. Comparison between MMMD–K-means–SDAE and MMMD–BPNN for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. MMMD–BPNN		
	30 min	1 h	2 h
P_{MRE} (%)	52.49	60.68	55.58
P_{RMSE} (%)	45.63	57.31	54.38
P_{MAPE} (%)	31.07	37.29	33.35

TABLE 20. Comparison between MMMD–K-means–SDAE and MMMD–SDAE for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. MMMD–SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	19.61	26.05	28.40
P_{RMSE} (%)	13.23	21.69	24.87
P_{MAPE} (%)	10.34	14.00	15.38

TABLE 21. Comparison between MMMD–K-means–SDAE and K-means–SDAE for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. K-means–SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	41.20	48.88	43.00
P_{RMSE} (%)	33.10	44.79	41.65
P_{MAPE} (%)	23.32	28.51	24.16

TABLE 22. Comparison between MMMD–K-means–SDAE and WT–K-means–SDAE for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. WT–K-means–SDAE		
	30 min	1 h	2 h
P_{MRE} (%)	2.81	18.19	25.67
P_{RMSE} (%)	2.23	15.57	24.09
P_{MAPE} (%)	1.41	9.55	13.79

were 62.89, 55.56, and 51.81, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model could improve P_{MRE} by 68.82, 74.42, and 68.43, respectively,

TABLE 23. Comparison between MMMD–K-means–SDAE and VMD–K-means–LSTM for wind turbine #2.

Indices \ Scales	MMMD–K-means–SDAE vs. VMD–K-means–LSTM		
	30 min	1 h	2 h
P_{MRE} (%)	2.55	17.31	25.62
P_{RMSE} (%)	1.32	14.01	23.37
P_{MAPE} (%)	1.28	9.06	13.76

TABLE 24. Further comparison of forecasting errors between the proposed and benchmark models for wind turbine #1.

Forecasting approaches	Time scale	P_{MRE} (%)	P_{RMSE} (%)	P_{MAPE} (%)
BPNN approach	30 min	86.22	85.20	62.89
	1 h	80.25	79.81	55.56
	2 h	76.78	77.16	51.81
SDAE approach	30 min	83.80	82.24	59.75
	1 h	74.43	74.29	49.43
	2 h	71.26	71.19	46.39
MMMD–BPNN approach	30 min	77.99	76.84	53.08
	1 h	68.96	69.11	44.28
	2 h	63.05	63.54	39.21
MMMD–SDAE approach	30 min	42.00	39.09	23.84
	1 h	28.59	27.91	15.49
	2 h	33.75	35.09	18.60
K-means–SDAE approach	30 min	68.51	66.52	43.84
	1 h	55.45	55.15	33.25
	2 h	51.43	51.68	30.31
WT–K-means–SDAE approach	30 min	14.20	12.29	7.38
	1 h	13.60	10.12	7.04
	2 h	23.86	15.35	12.74
VMD–K-means–LSTM approach	30 min	13.35	11.94	6.96
	1 h	12.95	9.62	6.69
	2 h	23.04	15.26	12.27

P_{RMSE} by 63.76, 71.86, and 67.25, respectively, and P_{MAPE} by 44.16, 49.41, and 43.81, respectively. The prediction results from the proposed MMMD–K-means–SDAE model were closer to the actual values.

6) MMMD–K-MEANS–SDAE VS. MMMD–BPNN

Compared to the MMMD–BPNN approach, the proposed hybrid MMMD–K-means–SDAE produced a better forecasting performance. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h. The improvements in P_{MRE} were 77.99, 68.96, and 63.05, respectively, whereas those in P_{RMSE} were 76.84, 69.11, and 63.54, respectively, and those in P_{MAPE} were 53.08, 44.28, and 39.21, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model could improve P_{MRE} by 52.49, 60.68, and 55.58, respectively, P_{RMSE} by 45.63, 57.31, and 54.38, respectively, and P_{MAPE}

TABLE 25. Further comparison of forecasting errors between the proposed and benchmark models for wind turbine #2.

Forecasting approaches	Time scale	P _{MRE} (%)	P _{RMSE} (%)	P _{MAPE} (%)
BPNN approach	30 min	68.82	63.76	44.16
	1 h	74.42	71.86	49.41
	2 h	68.43	67.25	43.81
SDAE approach	30 min	62.02	56.25	38.36
	1 h	69.97	66.70	45.19
	2 h	63.05	61.17	39.21
MMMD–BPNN approach	30 min	52.49	45.63	31.07
	1 h	60.68	57.31	37.29
	2 h	55.58	54.38	33.35
MMMD–SDAE approach	30 min	19.61	13.23	10.34
	1 h	26.05	21.69	14.00
	2 h	28.40	24.87	15.38
K-means–SDAE approach	30 min	41.20	33.10	23.32
	1 h	48.88	44.79	28.51
	2 h	43.00	41.65	24.16
WT-K-means-SDAE approach	30 min	2.81	2.23	1.41
	1 h	18.91	15.57	9.55
	2 h	25.67	24.09	13.79
VMD-K-means-LSTM approach	30 min	2.55	1.32	1.28
	1 h	17.31	14.01	9.06
	2 h	25.62	23.37	13.76

by 31.07, 37.29, and 33.35, respectively. The prediction results from the proposed MMMD–K-means–SDAE model were closer to the actual values.

7) MMMD–K-MEANS–SDAE VS. MMMD–SDAE

Compared to the MMMD–SDAE approach, the proposed hybrid MMMD–K-means–SDAE produced a better forecasting performance. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h. The improvements in P_{MRE} were 42.00, 28.59, and 33.75, respectively, whereas those in P_{RMSE} were 39.09, 27.91, and 35.09, respectively, and those in P_{MAPE} were 23.84, 15.49, and 18.60, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model could improve P_{MRE} by 19.61, 26.05, 28.40, respectively, whereas P_{RMSE} could improve by 13.23, 21.69, and 24.87, respectively, and P_{MAPE} by 10.34, 14.00, and 15.38, respectively. The prediction results from the proposed MMMD–K-means–SDAE model are closer to the actual value.

8) MMMD–K-MEANS–SDAE VS. K-MEANS–SDAE

Compared to the K-means–SDAE approach, the proposed hybrid MMMD–K-means–SDAE produced a better forecasting performance. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h, respectively. The improvements in P_{MRE} were 68.51, 55.45, and 51.43, respectively, whereas those in

P_{RMSE} were 66.52, 55.15, and 51.68 respectively, and those in P_{MAPE} were 43.84, 33.25, and 30.31, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model could improve P_{MRE} by 41.20, 48.88, and 43.00, respectively, whereas P_{RMSE} could improve by 33.10, 44.79, and 41.65, respectively, and P_{MAPE} by 23.32, 28.51, and 24.16, respectively. The prediction results from the proposed MMMD–K-means–SDAE model are closer to the actual values.

9) MMMD–K-MEANS–SDAE VS. WT-K-MEANS-SDAE MODELS

Compared to the WT-K-means-SDAE model, the proposed hybrid MMMD–K-means–SDAE produced a better forecasting performance. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h, respectively. The improvements in P_{MRE} were 14.20, 13.60, and 23.86, respectively, whereas those in P_{RMSE} were 12.29, 10.12, and 15.35 respectively, and those in P_{MAPE} were 7.38, 7.04, and 12.74, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model could improve P_{MRE} by 2.81, 18.19, and 25.67, respectively, P_{RMSE} by 2.23, 15.57, and 24.09, respectively, and P_{MAPE} by 1.41, 9.55, and 13.79, respectively. The prediction results from the proposed MMMD–K-means–SDAE model are closer to the actual values.

10) MMMD–K-MEANS–SDAE VS. VMD-K-MEANS-LSTM MODELS

Compared to the VMD-K-means-LSTM model, the proposed hybrid MMMD–K-means–SDAE produced a better forecasting performance. For wind turbine #1, the time scales were 30 min, 1 h, and 2 h, respectively. The improvements in P_{MRE} were 13.45, 12.92, and 23.04, respectively, whereas those in P_{RMSE} were 11.94, 9.62, and 15.26, respectively, and those in P_{MAPE} were 6.96, 6.69, and 12.27, respectively. For wind turbine #2 at the same time scales, the proposed hybrid model could improve P_{MRE} by 2.55, 17.31, and 25.62, respectively, P_{RMSE} by 1.32, 14.01, and 23.37, respectively, and P_{MAPE} by 1.28, 9.06, and 13.76, respectively. The prediction results from the proposed MMMD–K-means–SDAE model are closer to the actual values.

IV. CONCLUSION

This article proposed a hybrid MMMD–K-means–SDAE model for short-term wind-speed forecasting on multiple scales. In the proposed model, MMMD analysis and K-means clustering are used as signal-decomposing and data-mining algorithms, respectively, and SDAE is utilized as the core forecasting unit to capture the characteristics of each cluster. The predictions of all subsequences are then aggregated to obtain the final predicted wind speed. The complexity of the proposed model is constant, and it does not vary with the size of the datasets. However, when the amount of data increases, the operating cost of the model, such as the computing time and hardware cost, will increase; therefore, the total cost of the model will change with the size of the datasets.

Five different benchmark models and two novel hybrid wind forecasting models were implemented for comparison with our proposed model. First, the results show the superior ability of the SDAE in processing nonlinear and nonstationary wind-speed signals compared to the traditional BP network. Second, we use MMD technology to decompose wind-speed sequences into time-domain subsequences in order to avoid errors during the signal transfer process and significantly promote the reliability and precision of wind-speed forecasting. Third, the K-means algorithm, as a clustering analysis approach, can further enhance the prediction ability of the MMD–SDAE model. Finally, the main advantage of the proposed model is that its prediction accuracy is higher than that of five benchmarks models, particularly two novel hybrid models. Furthermore, we use MMD to decompose the original wind-speed sequence directly in the time domain; this is the first time a time domain signal-decomposing tool has been used in a hybrid wind-speed forecasting model.

However, many outliers exist in the data owing to the natural characteristics of wind speed. When the volatility of wind speed is high, the amount of outlier data will increase, and the accuracy of the prediction results as well as reliability of the entire system will be significantly reduced. Furthermore, the effective prediction time range is short (30 min to 2 h). Beyond this prediction range, the prediction sensitivity decreases with time. When the time scale exceeds 48 h, the model cannot provide a valid prediction. These two limitations are the main topics that will be addressed in future research.

REFERENCES

- [1] F. Blaabjerg and K. Ma, "Future on power electronics for wind turbine systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 1, no. 3, pp. 139–152, Sep. 2013.
- [2] M. Hossain, B. Rejabdar, S. J. Louis, and S. Dascalu, "Forecasting the weather of nevada: A deep learning approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–6.
- [3] C. Skitides and W.-G. Früh, "Wind forecasting using principal component analysis," *Renew. Energy*, vol. 69, pp. 365–374, Sep. 2014.
- [4] G. Q. Zhang and B. M. Zhang, "Wind speed and wind turbine output forecast based on combination method," *Autom. Electr. Power Syst.*, vol. 33, no. 18, pp. 92–96, Sep. 2009.
- [5] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3276–3284, May 2018.
- [6] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *Proc. North Amer. Power Symp.*, Sep. 2010, pp. 1–8.
- [7] M. Lei, L. Shiyang, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, May 2009.
- [8] Q. Xu, D. He, N. Zhang, C. Kang, Q. Xia, J. Bai, and J. Huang, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1283–1291, Oct. 2015.
- [9] M. G. Lobo and I. Sanchez, "Regional wind power forecasting based on smoothing techniques, with application to the spanish peninsular system," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1990–1997, Nov. 2012.
- [10] E. Yatiyana, S. Rajakaruna, and A. Ghosh, "Wind speed and direction forecasting for wind power generation using ARIMA model," in *Proc. Australas. Universities Power Eng. Conf. (AUPEC)*, Nov. 2017, pp. 1–6.
- [11] I. Colak, S. Sagirolu, M. Yesilbudak, E. Kabalci, and H. I. Bulbul, "Multi-time series and -time scale modeling for wind speed and wind power forecasting—Part I: Statistical methods, very short-term and short-term applications," in *Proc. Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, Nov. 2015, pp. 209–214.
- [12] P. Xiaosheng, W. Bo, Y. Fan, F. Gaofeng, W. Zheng, and C. Kai, "A deep learning approach for wind power prediction based on stacked denoising auto encoders optimized by bat algorithm," in *Proc. China Int. Conf. Electr. Distrib. (CICED)*, Sep. 2018, pp. 945–948.
- [13] Y. Deng, H. Jia, P. Li, X. Tong, X. Qiu, and F. Li, "A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction," in *Proc. 14th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2019, pp. 591–595.
- [14] X. Wang, Y. Yang, and C. Li, "Deep belief network based multi-dimensional phase space for short-term wind speed forecasting," in *Proc. Int. Conf. Sens., Diagnostics, Prognostics, Control (SDPC)*, Aug. 2018, pp. 204–208.
- [15] K.-P. Lin, P.-F. Pai, and Y.-J. Ting, "Deep belief networks with genetic algorithms in forecasting wind speed," *IEEE Access*, vol. 7, pp. 99244–99253, 2019.
- [16] Z. Sun, S. Zhao, and J. Zhang, "Short-term wind power forecasting on multiple scales using VMD decomposition, K-means clustering and LSTM principal computing," *IEEE Access*, vol. 7, pp. 166917–166929, 2019.
- [17] C. Yu, Y. Li, and M. Zhang, "An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on elman neural network," *Energy Convers. Manage.*, vol. 148, pp. 895–904, Sep. 2017.
- [18] L. Wang, L. Dong, Y. Hao, and X. Liao, "Wind power prediction using wavelet transform and chaotic characteristics," in *Proc. World Non-Grid-Connected Wind Power Energy Conf.*, Sep. 2009, pp. 1–5.
- [19] R. Ye, Z. Guo, and R. Liu, "Wind speed and wind power forecasting method based on WPD and improved Elman neural network," *Trans. China Electrotech. Soc.*, vol. 32, no. 21, pp. 103–111, Jul. 2017.
- [20] P. Chen, H. Y. Chen, and R. Ye, "Wind speed forecasting based on multi-scale morphological analysis," *Power Syst. Protection Control.*, vol. 38, no. 21, pp. 12–18, Nov. 2010.
- [21] L. Zhang, J. Xu, J. Yang, D. Yang, and D. Wang, "Multiscale morphology analysis and its application to fault diagnosis," *Mech. Syst. Signal Process.*, vol. 22, no. 3, pp. 597–610, Apr. 2008.
- [22] Y. F. Wan, J. Jiang, Z. Y. Zhen, and P. Zhu, "Multiscale morphology based SVD noise reduction method," *Electron. Opt. Control*, vol. 27, no. 1, pp. 21–25, Jan. 2020.
- [23] J. A. Bangham, P. Chardaire, C. J. Pye, and P. D. Ling, "Multiscale nonlinear decomposition: The sieve decomposition theorem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 5, pp. 529–539, May 1996.
- [24] R. M. Esteves, T. Hacker, and C. Rong, "Competitive K-means, a new accurate and distributed K-means algorithm for large datasets," in *Proc. IEEE 5th Int. Conf. Cloud Comput. Technol. Sci.*, Dec. 2013, pp. 17–24.
- [25] L. Yang and M. Deng, "Based on k-Means and fuzzy k-Means algorithm classification of precipitation," in *Proc. Int. Symp. Comput. Intell. Design*, Oct. 2010, pp. 211–218.
- [26] Y. Zhang, Z. Xiong, J. Mao, and L. Ou, "The study of parallel K-means algorithm," in *Proc. 6th World Congr. Intell. Control Autom.*, Jun. 2006, pp. 5868–5871.
- [27] F. Yi and I. Moon, "Extended K-means algorithm," in *Proc. 5th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, Aug. 2013, pp. 263–266.
- [28] Z. Sun and H. Sun, "Stacked denoising autoencoder with density-grid based clustering method for detecting outlier of wind turbine components," *IEEE Access*, vol. 7, pp. 13078–13091, 2019.
- [29] Z. Li and X. Dang, "A stock forecasting method based on combination of SDAE and BP," in *Proc. Int. Conf. Orange Technol. (ICOT)*, Oct. 2018, pp. 1–6.
- [30] E. K. Wang, X. Zhang, and L. Y. Pan, "Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term network," *IEEE Access*, vol. 7, pp. 182873–182880, 2019.
- [31] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 9, pp. 2391–2402, Sep. 2017.



WEICHAO DONG was born in Shijiazhuang, Hebei, China, in 1989. He received the B.Sc. degree in electrical engineering and the M.Sc. degree in electrical and computing engineering from Cornell University, NY, USA, in 2012 and 2013, respectively. He is currently pursuing the Ph.D. degree in control theory and control engineering with the Hebei University of Technology, Tianjin, China. He is a Lecturer with the College of Electrical Engineering, Hebei University of Science and Technology, China. He has published two articles on SCI. His research interests include optimization of wind engine structure, wind power, and the application of deep learning in wind energy.



JINGXUAN ZHANG (Graduate Student Member, IEEE) was born in Hebei, China, in 1982. He received the B.Sc. degree in electrical automation and the M.Sc. degree in control science and engineering from the Hebei University of Technology, Tianjin, China, in 2007 and 2010, respectively, where he is currently pursuing the Ph.D. degree in control theory and control engineering. He is also currently a Lecturer with the College of Electrical Engineering, North China University of Science and Technology, China. He has published two articles on SCI and EI journals. His research interests include fault diagnosis, switched systems, and power systems. He is a PES Member.



HEXU SUN (Senior Member, IEEE) received the Ph.D. degree in automation from Northeastern University, Shenyang, China, in 1993.

He has been a Professor with the School of Control Science and Engineering, Hebei University of Technology, Tianjin, China, and the School of Electrical Engineering, Hebei University of Science and Technology, Shijiazhuang, China. He has authored five books and more than 130 journal and conference papers. He holds 13 U.S. patents and

five computer software copyrights. His current research interests include robotics and complex engineering systems.

Dr. Sun was a recipient of many prestigious national awards from China. He has been the Director in many societies and committees in China. He is currently the invited Plenary Speaker and the General Co-Chair of many international conferences.



ZHENG LI (Member, IEEE) was born in Shijiazhuang, Hebei, China, in 1980. He received the B.Sc. and Ph.D. degrees in electrical engineering and power electronics and electric drive from the Hefei University of Technology, Hefei, China, in 2002 and 2007, respectively.

Since 2007, he has been a Lecturer, an Associate Professor, and a Professor with the School of Electrical Engineering, Hebei University of Science and Technology. From July 2013 to July 2014,

he was a Visiting Scholar and a part-time Faculty with the College of Engineering, Wayne State University, USA. He is the author of more than 160 published articles. His current research interests include design, analysis, and control of novel motors and actuators, intelligent control, and power electronic.

Dr. Li is an Active Reviewer for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON ENERGY CONVERSION, the IEEE TRANSACTIONS ON MAGNETICS AND ELECTRIC POWER COMPONENTS AND SYSTEMS, and so on.



HUIFANG YANG was born in Shijiazhuang, Hebei, China, in 1964. She received the Ph.D. degree, in 2007.

She has been a Professor with the School of Chemical Engineering, Shijiazhuang Tiedao University, since 2002. In 2012, she was a Visiting Scholar and a part-time Faculty with the College of Chemical Engineering, University of Delaware, USA. She has authored more than 30 journal and conference papers. Her current research interests

include nano material, materials science, and application of materials science in new energy sources. She was a recipient of many prestigious national awards from China.

...