Applied Energy 249 (2019) 392–408

Contents lists available at ScienceDirect

Applied Energy

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

Data processing strategies in wind energy forecasting models and applications: A comprehensive review

Hui Liu*, Chao Chen

Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

HIGHLIGHTS

- Data processing methods for wind energy forecasting models are reviewed.
- Seven categories are classified based on their computing procedures.
- A general evaluation from different perspectives is presented for the reviewing.
- Trends and challenges in the application of data processing methods are discussed.

ARTICLE INFO

Keywords:
Wind energy forecasting
Data processing
Decomposition
Dimensionality reduction
Data correction

ABSTRACT

Given the intermittent nature of the wind, accurate wind energy forecasting is significant to the proper utilization of renewable energy sources. In recent years, data-driven models based on past observations have been widely employed in the literature. Various types of data processing methods are successfully applied to assist these models and further improve forecasting performance. Comprehensive research of their methodologies is called on for a thorough understanding of current challenges that affect model accuracy and efficiency. To address the knowledge gap, this paper presents an exhaustive review and categorization of data processing in wind energy forecasting. The utilized techniques are classified into seven categories according to the applications: decomposition, feature selection, feature extraction, denoising, residual error modeling, outlier detection, and filter-based correction. An overall analysis of their intentions, positions, characteristics, and implementation details is provided. A general evaluation is carried out from different perspectives including accuracy improvement, usage frequency, consuming time, robustness to parameters, maturity, and implementation difficulty. Among the existing data processing methods, outlier detection and filter-based correction are relatively less used. Their potential can be better explored in the future. Furthermore, some possible research directions and challenges of data processing in wind energy forecasting are provided, in order to encourage subsequent research.

1. Introduction

Wind energy is currently the most promising renewable energy source, which has gained increasing attention from all over the world. Wind power is conducive to the protection of global environmental resources. However, the wind resources are uncontrollable, and the internal instability makes the utilization rate relatively low [1]. Accurate prediction of wind power and wind speed has been attached great significance to the exploiting of renewable energy sources. Many researchers have made great efforts in the development and perfection of wind speed and wind power forecasting.

There are two basic criteria for the preliminary classification of wind energy forecasting models: prediction time horizon and forecasting approach. With respect to time horizon, wind energy forecasting can be divided into four categories, including very short-term, short-term, medium-term, and long-term [2,3]. The time-scale classification of forecasting models is relatively vague, as shown in Fig. 1. In general, shorter forecasting time horizon can provide more detailed and accurate results, but less time left for the deployment of wind power generation. Longer forecasting time horizon provides long-term information about future wind energy, but usually, the accuracy is relatively

* Corresponding author at: Institute of Artificial Intelligence and Robotics, Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, Hunan, China.

E-mail address: csuliuhui@csu.edu.cn (H. Liu).

https://doi.org/10.1016/j.apenergy.2019.04.188
Received 2 March 2019; Received in revised form 27 April 2019; Accepted 30 April 2019
Available online 09 May 2019
0306-2619/

Nomenclature

List of abbreviations

ACF	autocorrelation function
ANN	artificial neural network
ARIMA	autoregressive integrated moving average model
BBSA	binary backtracking search algorithm
BPNN	back propagation neural network
BPSOGSA	binary particle swarm optimization and gravitational search algorithm
CEEMD	complementary empirical mode decomposition
CEEMDAN	complete ensemble empirical mode decomposition with adaptive noise
CMI	conditional mutual information
CRO	coral reefs optimization algorithm
DBSCAN	density-based spatial clustering of applications with noise
DE	differential evolution
DWT	discrete wavelet transform
EEMD	ensemble empirical mode decomposition
EMD	empirical mode decomposition
EWT	empirical wavelet transform
FEEMD	fast ensemble empirical mode decomposition
GA	genetic algorithm
GARCH	generalized autoregressive conditional heteroskedasticity
GCA	grey correlation analysis
GCT	granger causality test
GSO	Gram–Schmidt orthogonal
ICEEMDAN	improved CEEMDAN
IEWT	inverse empirical wavelet transform
IWT	inverse wavelet transform
KFCM	kernel-based fuzzy c-means clustering
KLD	Kullback-Leibler divergence
KPCA	kernel principal component analysis
LOF	local outlier factor
LSSVM	least squares support vector machine
LSTM	long short-term memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MI	mutual information
MODWT	maximum overlap discrete wavelet transform
NP	non-parametric
OVMD	optimized variational mode decomposition
PACF	partial autocorrelation function
PCA	principal component analysis
PSO	particle swarm optimization
PSR	phase space reconstruction
RELM	regularized extreme learning machine
RF	random forest
RF	random forest
RFE	recursive feature elimination
RMSE	root mean square error
SC	spectral clustering
SE	sample entropy
SSA	singular spectrum analysis
SVD	singular value decomposition
SVM	support vector machine
TVF-EMD	time varying filter based empirical mode decomposition
VMD	variational mode decomposition
WF	wavelet filter
WPD	wavelet packet decomposition
WT	wavelet transform
WTD	wavelet threshold denoising

poor. The basic role of wind energy forecasting models is to provide the wind speed and power status for the next few seconds, minutes, hours or longer. Models with different forecast lead time have specific applications, as presented in Table 1.

According to forecasting approaches, four types of mainstream models have been proposed, known as physical, statistical, intelligent and hybrid models. Physical models mainly refer to numerical weather prediction (NWP) which utilizes various meteorological data collected from observing systems to simulate the trend of wind speed. They have limited practical utility in the 0–2 h forecast lead time due to latency issues, and need accurate initial conditions of wind farms that cannot be always guaranteed [4]. However, with significant weather situations, the NWP models can represent weather phenomena that can cause wind ramps, such as fronts and thunderstorms. Different from them, the core concept of statistical and intelligent models is to regard the forecast of wind energy as a stochastic process [5] and use past observations to mine the time-varying relationships within time series. The most popular statistical methods include the Kalman filter [6], Box. Jenkins models (AR, ARIMA models, etc.) [7], Bayesian regression [8], and so on. They can achieve satisfactory accuracy in simple time series forecasting. But their applications are limited because of inadequate ability to process nonlinear data. Associated with machine learning, the intelligent models are increasingly prevailing for wind energy forecasting, such as multilayer perceptron (MLP) [9], support vector machine (SVM) [10], and recurrent neural network(RNN) [11]. These models generally outperform traditional statistical methods owing to the advantages of self-organization and adaptive learning.

Several methods have been presented based on a single forecasting algorithm, such as convolutional neural network (CNN) [12], echo state network (ESN) [13] and deep neural network (DNN) [14], whose effectiveness has been demonstrated by comparison experiments. Nevertheless, the single approaches cannot always guarantee high accuracy due to some inherent disadvantages, such as lacking theoretical standards for determining parameters and insufficient robustness to noise. For this purpose, various methods have been applied in hybrid models to obtain more accurate and robust forecasts. For instance, the NWP models postprocessed by statistical methods can generate a better forecast than either method alone [15]. Moreover, to solve the parameter setting of models, metaheuristic optimization algorithms are employed to adaptively search for globally optimal solutions. Examples include differential evolution (DE) [16], grey wolf optimization (GWO) [17], cuckoo search (CS) [18], multi-objective whale optimization algorithm (MOWOA) [19], and etc. Among them, multi-objective optimization algorithms have been proved to be effective to synchronously guarantee model accuracy and stability [20,21]. To overcome the instability and weakness of a single model, ensemble learning integrates multiple base models and generates a stronger predictor. Some famous ensemble methods have been adopted to ensure the diversity of base

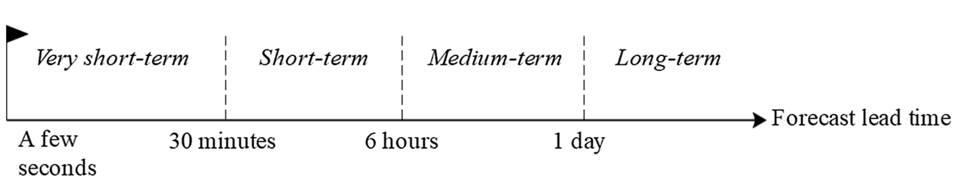

Fig. 1. Time-scale classification of wind energy forecasting.

Table 1
Applications of wind energy forecasting models with different time horizons.

Time horizon	Applications
Very short-term	– Turbine regulation
	– Electricity market clearing
Short-term	– Preload sharing
Medium-term	– Energy allocation
	– Electric power system management
Long-term	– Maintenance scheduling of related equipment

predictors, such as adaptive boosting (AdaBoost) [22] and bootstrap aggregation (Bagging) [23]. In addition to the abovementioned two types of methods, many data processing methods have been widely utilized to assist the data-driven models. The decomposition of wind speed series, for example, can make the input data more stationary. Trained by the preprocessed data, the data-driven models are more likely to learn the nonlinear characteristics of the wind. The effectiveness of combining data processing has been verified in the aspect of both accuracy and efficiency.

Given the significance of wind energy forecasting, many review papers have been focused on the methodologies in it. Yan et al. reviewed the uncertainty analysis of wind power forecasting [24]. Okumus et al. encapsulated the current status of wind energy forecasting and proposed a hybrid method for hourly predictions [25]. Marugán surveyed the applications of ANN in wind energy systems [26]. Wang et al. summarized eight multi-step-ahead forecasting strategies and compared them on intelligent models [27]. Shi et al. evaluated the hybrid forecasting approaches of ANN, SVM and ARIMA [28]. Tascikaraoglu et al. reviewed the hybrid wind energy forecasting models [29]. Zendehboudi et al. surveyed the applications of SVM in solar and wind energy forecasting [30]. These reviews summarized the wind energy forecasting models from different perspectives. Although some of them mentioned about data processing, the induction and classification are relatively not in-depth and blanket. For instance, decomposition is not theonly way to preprocess the original wind energy data. Preprocessing methods like feature selection and denoising haven't drawn much attention. But they are widely used, and their effectiveness has been verified in many research papers. A comprehensive review of these methods is called on. The main contribution of this paper is to present a clear categorization of the applications of data processing in wind energy forecasting, as well as the theory and implementation details. Researchers in wind energy forecasting can get knowledge of the current status, future trends and challenges. Besides, for the less commonly used methods, we hope they can attract more attention and be widely experimented in the future.

The structure of the paper is organized as follows: Section 2 gives a general classification and introduction of the discussed data processing technologies. Sections 3–5 review the data processing methods which are adopted for different purposes: data decomposition, dimensionality reduction, and data correction. Section 6 presents a comprehensive discussion of these methods, including performance evaluation, future research directions, and challenges. Finally, Section 7 concludes the paper.

2. Data processing applications in wind energy forecasting

A graphical summary of the technologies discussed in this paper is shown in Fig. 2. Seven categories of data processing methods have been reported in hybrid intelligent models: (1) decomposition, (2) feature

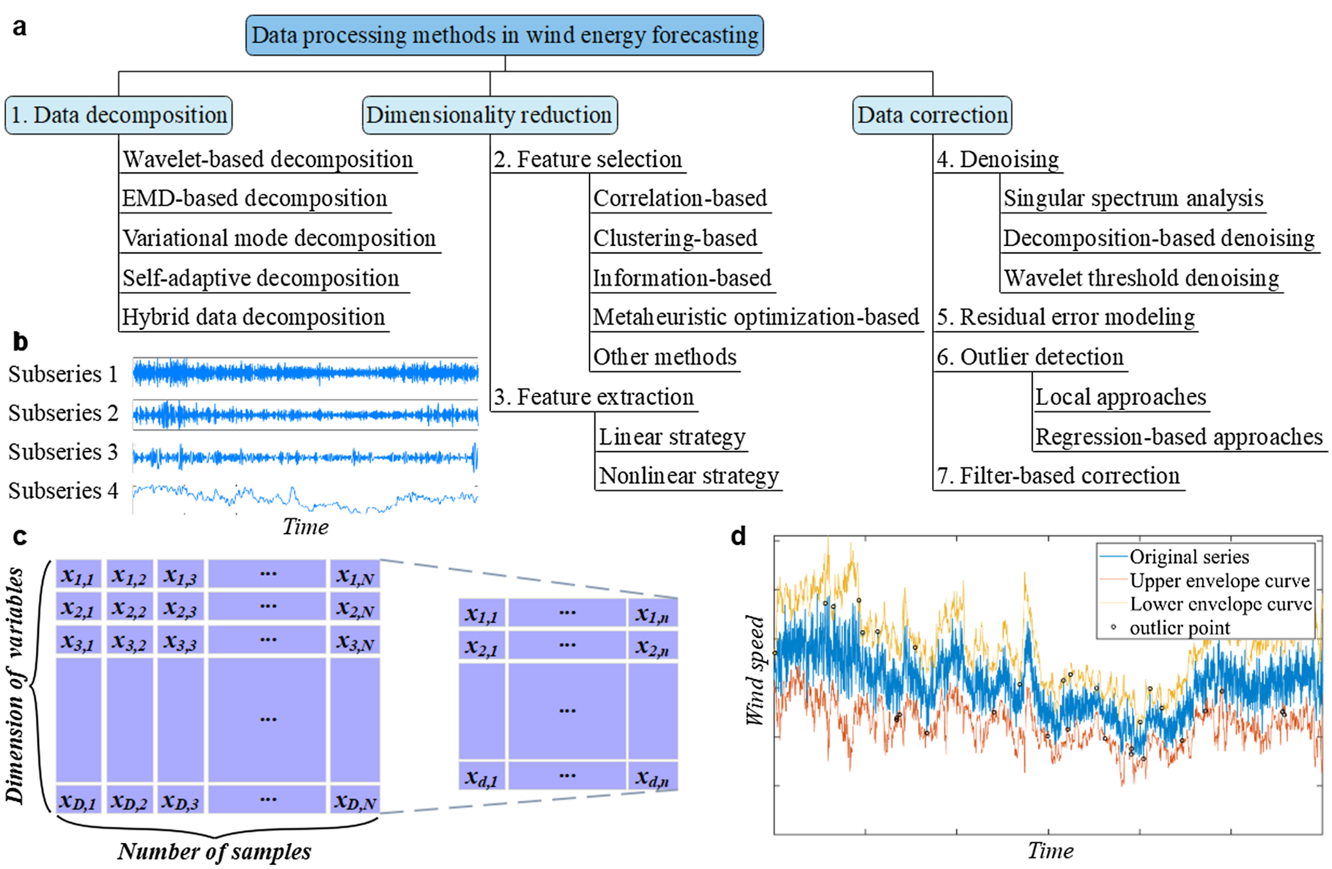

Fig. 2. A graphical summary of the technologies discussed in this paper. (a) Classification of data processing methods in wind energy forecasting. (b) Results of data decomposition: a set of relatively stationary subseries. In the example, the subseries 1–3 have high frequencies, while the subseries 4 represents the major trend component of original wind speed series. (c) The purpose of feature selection: original *D*-dimensional variables and *N*-dimensional samples are reduced to smaller subsets. (d) A case of data correction. The data points beyond the bounds of the envelope will be corrected.

selection, (3) feature extraction, (4) denoising, (5) residual error modeling, (6) outlier detection and (7) filter-based correction. They are adopted in forecasting models for different purposes. In general, decomposition is adopted to turn the original series into several relatively stationary subseries, so as to reduce forecasting difficulty. Feature selection and feature extraction are used to reduce dimensionality. And the other four methods are employed to correct input data sets or forecasting results.

Concerning their positions in the model, the adopted methods are integrated with different strategies, namely preprocessing and postprocessing. They have different intentions and processed results, as shown in Table 2. In general, the main concept of preprocessing is to reduce the complexity or noise of original data before being fed into predictive models, while the postprocessing is to enhance preliminary forecasting output.

3. Data decomposition

In recent years, more than 100 research papers have focused on decomposition-based hybrid models [31]. In the published literature, the decomposition-based hybrid models basically adopt the same framework, which can be named as the decomposition-and-integration framework, as shown in Fig. 3. The framework utilizes decomposition methods to decompose non-stationary original time series into several relatively stationary subseries, and then builds a forecasting model on each subseries to get several individual forecasting results. The final forecasting result is obtained by adding up all individual forecasting results. By independently forecasting each subseries using the predictor can improve the forecasting accuracy of the wind energy series. The effectiveness of decomposition can also be explained from the aspect of frequency. The subseries have with more concentrated frequency bands, which makes the predictors only need to focus on the components of a single frequency band and reduce the forecasting difficulty. After a comprehensive literature review, the mainstream decomposition methods used in wind energy forecasting can be further divided into five categories: (a) wavelet-based, (b) EMD-based, (c) variational mode decomposition, (d) self-adaptive decomposition and (e) hybrid decomposition.

3.1. Wavelet-based decomposition

Wavelet analysis theory is the most outstanding representative achievement in the field of intelligent control and applied mathematics. It has both good localization properties in the time domain and frequency domain, which can decompose the signal into components in different frequency bands, and then obtain high-resolution decomposition results.

Wang et al. performed wavelet transform (WT) based on the Daubechies wavelet with nine vanishing moments and three decomposition layers on the original wind energy series to generate an approximate series **A3**, and three detailed series, **D1**, **D2** and **D3** [32]. In practical applications, continuous wavelet and continuous wavelet transform must be processed discretely for computer operation, which results in the discrete wavelet transform (DWT). Nevertheless, the traditional DWT has a limitation on the size of the input signal. In order to overcome this problem, another wavelet-based decomposition method named maximum overlap discrete wavelet transform (MODWT) is also utilized to preprocess the wind speed series [33]. The use of MODWT and multiresolution analysis is quite helpful in wind speed forecasting.

The WT only subdivides the approximate components of each layer. However, in many applications, the time series contains a lot of information in the intermediate frequency or high frequency. It is

Table 2
The intentions, results, and positions of classified seven data processing methods.

Category	Intentions	Results	Preprocessing	Postprocessing
Decomposition	● Reduce forecasting difficulty	Several relatively stationary subseries	√	
Feature selection	● Remove redundant features or get the minimum subset	Subset of original sets	√	
Feature extraction	● Reduce the dimension with only a small loss of information	New features	√	
Denoising	● Sort out and strengthen the effective information of input data	Denoised series	√	
Residual error modeling	● Correct the systematic errors	Corrected series		√
Outlier detection	● Improve data set quality or correct forecasting result	Data without outliers	√	√
Filter-based correction	● Remove the extra components generated by forecasting	Corrected series		√

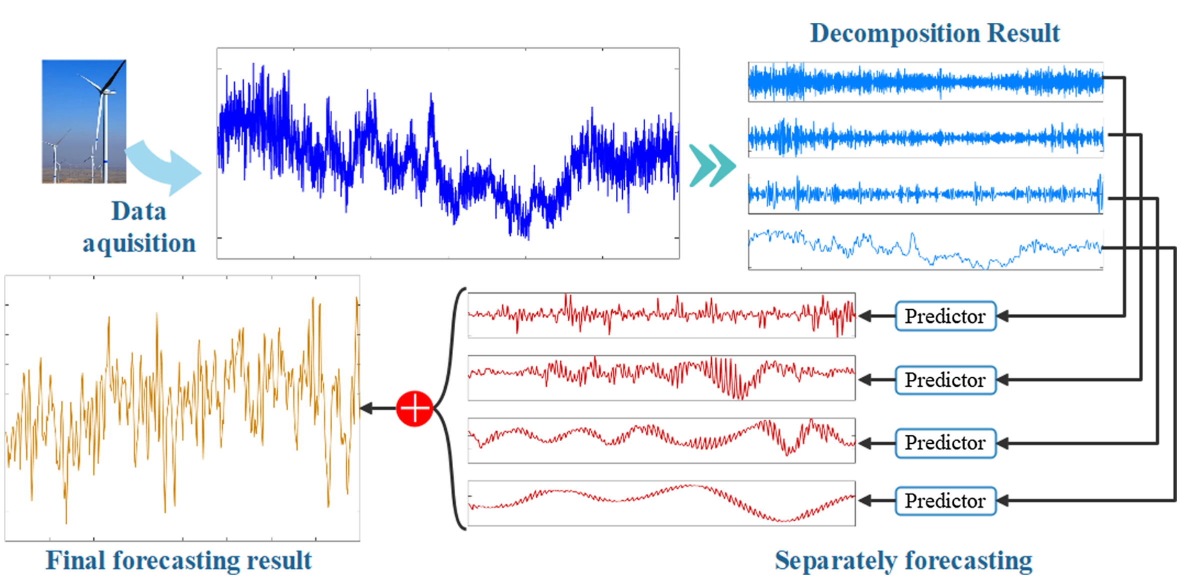

Fig. 3. The decomposition-and-integration framework of wind energy forecasting.

necessary to further decompose the high-frequency detailed components. Wavelet packet decomposition (WPD) is a variant of wavelet decomposition. The binary trees of WT and WPD with three decomposition layers are shown in Fig. 4. WPD can decompose the approximation component (A) and detailed component (D) of each layer to obtain a complete decomposition binary tree and a more comprehensive information division. It has been proved that the WPD-ELM model outperforms the WT-ELM model because the WPD algorithm can get more information from the original wind speed data than WT [34]. Therefore, WPD is a better choice than WT when considering the decomposition of wind energy series by wavelet-based technologies.

3.2. EMD-based decomposition

Since the wavelet-based decomposition method greatly depends on the structure of the decomposition binary tree and the choice of wavelet function and decomposition level, sufficient experimental data or prior knowledge is needed to achieve good results in practical applications. To some extent, the non-adaptive nature limits its usage. Empirical mode decomposition (EMD) proposed by Huang et al. in the 1990s can decompose complex time series adaptively into a limited number of intrinsic mode functions (IMFs) [35]. The obtained subseries have an instantaneous frequency with a clear physical meaning, which can well describe the oscillation of the original time series at each local.

Many researchers employed the EMD to preprocess the wind energy series and the results proved the effectiveness [36]. In recent years, EMD has developed rapidly and generated many variations, which is shown in Fig. 5. The EMD is an adaptive signal decomposition method, and there are a few hyperparameters that need to be adjusted. But two main problems still exist in its use: (a) endpoint effect and (b) mode mixing. In the algorithm, it is necessary to use the cubic spline curve to fit the extreme points of the signal to construct the upper and lower envelopes. In the actual signal, the endpoint value is not necessarily an extreme value. If the endpoint value is treated as an extreme value, it will result in an endpoint effect. The mode mixing problem refers to signals with different frequency segments in an IMF, or signals in the same frequency segment appear in different IMFs. To solve this problem, a time varying filter based-EMD (TVF-EMD) was applied to decompose the wind speed series, which could make a further definition of IMF and adjust the cutoff frequency [37].

Another improved version of EMD is ensemble empirical mode decomposition (EEMD). The EEMD adds white noise to the original signal and evenly distributed over the entire time-frequency space of the signal. Since white noise has a uniformly distributed spectrum, signals of different scales are automatically associated with appropriate reference scales for white noise. Due to the characteristics of zero-mean noise, after multiple calculations, the noise will cancel out interactively, so the average of multiple sets ofeigenmode functions can be calculated as the final real decomposition result. Experimental results demonstrated that EEMD could help to determine the characteristics of wind speed series, so as to effectively improve the performance and robustness of forecasting models [38].

Because of the addition of auxiliary white noise in EEMD, each IMF has more noise in the reconstruction process. In order to eliminate residual noise, it is usually necessary to increase the number of EMD operations and perform a large overall averaging, which greatly reduce the computational efficiency of EEMD. However, when complementary ensemble empirical mode decomposition (CEEMD) adds noise, the opposite polarity is used, that is, white noise with the same amplitude but opposite phase is added to the target signal respectively, and the effect of eliminating residual noise can be quickly achieved by using only a small number of average times. Several literatures used CEEMD to decompose wind speed series and achieve better performance [39].

To achieve a lower computational cost, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) was adopted to preprocess the wind speed series, which has the following advantages: (1) Introduce additional noise coefficient vector to control the noise level of each decomposition stage; (2) Reconstruction is completed without noise; (3) The number of trials is less than EEMD and CEEMD; (4) The problem that the EEMD and CEEMD have a tendency to produce incorrect components is resolved [40,41]. However, the CEEMDAN still has two problems that need to be improved: (1) The mode contains a certain amount of residual noise; (2) Signal information appears “later” than EEMD, and there are some spurious modes at the beginning of the decomposition [42]. To address these problems, improved CEEMDAN (ICEEMDAN) was proposed and utilized to decompose wind speed series in Ref. [43]. Experimental results showed that comparing ICEEMDAN with EMD, EEMD and CEEMDAN, the average MAPE improvements were higher than 5.82% in the proposed models. Therefore, the conclusion can be made that the ICEEMDAN method is more powerful than EMD, EEMD and CEEMDAN in performing decomposition of wind energy series.

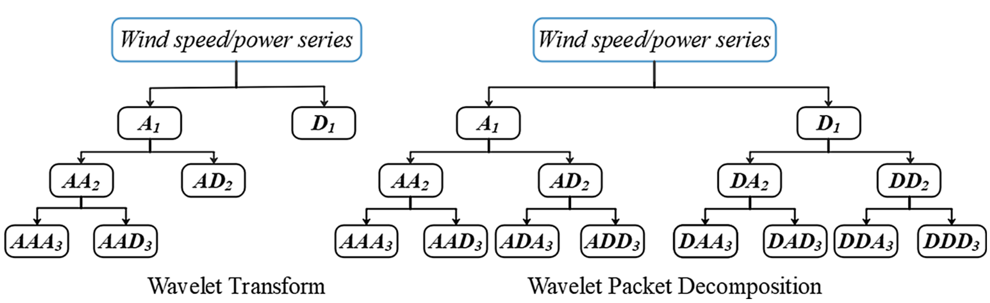

Fig. 4. The binary trees of WT and WPD. ***A*** represents the approximate subseries, while ***D*** represents the detailed subseries.

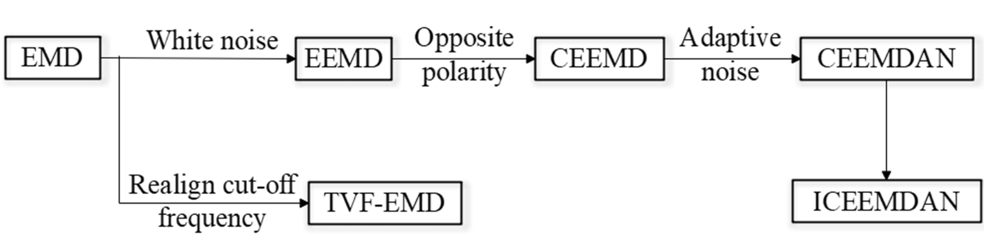

Fig. 5. The development and improvement process of the EMD and its variants.

3.3. Variational mode decomposition

Another advanced decomposition technology has been prevailing in recent years, namely variational mode decomposition (VMD), proposed by Dragomiretskiy and Zosso [44]. It can decompose the non-stationary signal into multiple band-limited intrinsic mode functions (BLIMFs) by non-recursive methods. Compared to EMD-based decomposition methods, VMD avoids errors caused during the calculation of recursion and the end of recursion. It has been widely applied in wind speed forecasting [45–48]. X. Wang et al. compared EEMD-based models with VMD-based models for short-term wind speed forecasting. Results showed that the former was sensitive to both noise and sampling. Oppositely, the VMD-based model demonstrated better noise robustness and more precise component separation [49].

3.4. Self-adaptive decomposition

All of the decomposition methods mentioned above need to predetermine several parameters, such as the wavelet function and decomposition level in WPD, the number of IMFs in EMD, and the number of modes in VMD. But the self-adaptive decomposition methods can adjust parameters adaptively. Two self-adaptive decomposition methods have been reported and applied in wind energy forecasting, namely optimized variational mode decomposition (OVMD) and empirical wavelet transform (EWT).

Dragomiretskiy reported that too few modes in VMD could lead to insufficient segmentation, while too many modes may capture additional noise or cause mode mixing problems [44]. In Ref. [50], C. Zhang et al. proposed a novel method to decide the optimal parameters of VMD. It firstly calculates the central frequency of decomposition mode with different k values. Once the phenomenon of similar frequency occurs, k-1 is selected as the optimal K for decomposition. Then, different values of the update parameter τ are tested and the best parameter is selected according to the root mean square error (RMSE) between the denoising time series and the original sequence. Hence, the number of modes in VMD can be adaptively determined.

The EWT [51] overcomes the defect of predetermining parameters, which can divide the frequency band autonomously according to the spectrum of the signal and generate a series of filters to decompose the data, as is shown in Fig. 6. Due to its excellent adaptiveness, many published papers have utilized EWT to decompose the original wind energy series [52]. In Ref. [53], three decomposition methods including EMD, WPD and EWT were utilized to decompose the wind speed series into several subseries. The low-frequency and high-frequency subseries were forecasted by long short-term memory neural network and Elman neural network, respectively. The performances of three different decomposition methods were compared. And the result showed that the EWT-based forecasting models had higher accuracy than models based on WPD and EMD.

3.5. Hybrid data decomposition

In order to further improve the performance of the decomposition method, some scholars proposed hybrid decomposition algorithms, also named as secondary decomposition algorithm (SDA) [54]. Generally, the hybrid decomposition algorithms further decompose the components obtained from the first-layer decomposition to extract more detailed components of the original sequence. Under the condition of reasonable decomposition scheme, the effect of the prediction model is better than that of other decomposition methods.

The selection of decomposed components is an important factor affecting the performance of hybrid decomposition technologies. Liu et al. [55] studied three different cases based on extreme learning machine (ELM): secondary decomposition of low-frequency, high-frequency or all components in the results of the first-layer decomposition. The comparison results showed that the second-order decomposition of low-frequency and all components could significantly improve the prediction accuracy, while the second-order decomposition of high frequency could not. It implies that determining which subseries of the first-layer decomposition result should be further decomposed is an important step in hybrid decomposition methods. In the literature, two strategies are adopted to solve this problem. They are summarized as follows:

Hybrid decomposition strategy 1: specify the subseries to be further decomposed. This hybrid decomposition strategy directly selects specific subseries to achieve secondary decomposition. The selected subseries is usually the one with the highest frequency. Yin et al. believed that the first intrinsic mode function (IMF1) derived from EMD contained a lot of potential characteristic information and found that utilizing WPD to process the IMF1 could significantly improve the accuracy of the model. An interesting phenomenon was that utilizing WPD to only process the IMF1 was better than the prediction results of utilizing WPD to process IMF1 and IMF2 at the same time [56]. Likewise, Peng et al. used VMD to decompose the component with the highest frequency obtained by the CEEMDAN. The higher the frequency was, the stronger the non-stationary and mutagenicity of the sequence, leading to the decrease of predictability [57]. Yu et al. employed WPD to decompose the wind speed series, and applied SSA to remove redundant information of the subseries with the highest frequency **D1** [58]. In Ref. [59], three-phase signal decomposition was proposed, which first obtained the seasonal and trend components by SSA. And then the trend component was further decomposed by a secondary decomposition method.

Hybrid decomposition strategy 2: use indicators to dynamically determine the subseries to be further decomposed. This hybrid decomposition strategy utilizes evaluation criteria to compare the stability or disorder of first-layer subseries. And the most non-stationary subseries are further decomposed. Sampleentropy (SE) is a commonly used

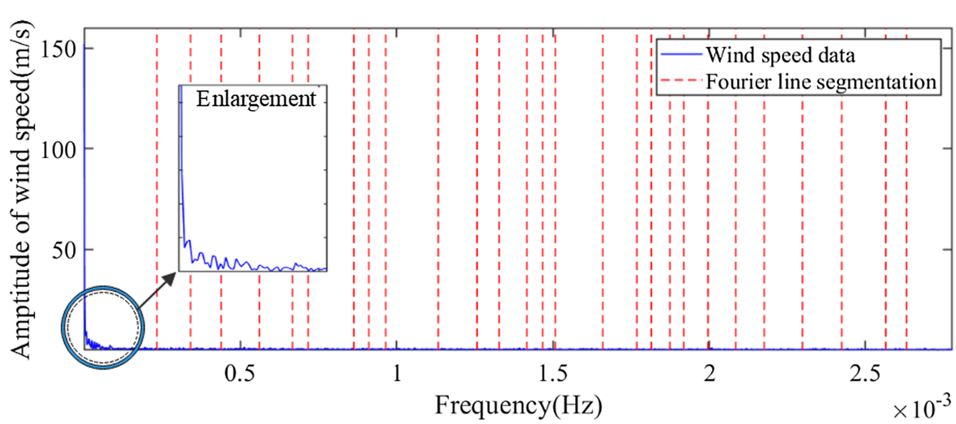

Fig. 6. EWT: the automatic detection of boundaries in the frequency domain. In this case, the wind speed series is decomposed into 28 segments.

evaluation criterion in hybrid decomposition technologies, which can well measure the predictability of sequences. The component with the highest SE generally demonstrates the most disordered subseries. Zhang et al. utilized the SE to estimate the complexity of subseries in the first-layer decomposition obtained by VMD and the subseries with the highest SE was further decomposed by WPD. Not surprisingly, the hybrid decomposition algorithm-based model obtained better accuracy than the other decomposition-based models [60]. Similarly, some other literatures employed the SE, such as CEEMD-SE-VMD [61] and WPD-SE-VMD [62]. Mi et al. adopted the augmented Dickey-Fuller test (ADF) to determine whether the subseries obtained by WPD were stationary [63]. The non-stationary subseries were further decomposed into a series of IMFs by EMD.

Both of the above hybrid decomposition strategies have been proved effective. The hybrid decomposition algorithm-based models generally perform better in the aspect of accuracy and stability. But as the number of subseries increases, it brings about the increase in computation time. When actually applying secondary decomposition algorithms, tradeoff should be made to balance the calculation accuracy and time.

Table 3 summarizes the abovementioned decomposition methods, including their characteristics, implementation details and utilized algorithms in references.

4. Dimensionality reduction

Apart from improving the forecasting accuracy by decomposing the original wind energy series, the management of input data is also concerned as a potential way to improve the performance of hybrid models. The dimensionality reduction of input data can greatly reduce computational complexity and the forecasting error caused by redundant information. Feature selection and feature extraction are two main ways to achieve dimensionality reduction. But they have some differences. Feature selection refers to selecting a subset from the existing feature set according to certain criteria, while the features obtained by feature extraction is a mapping of the original feature set, that is, feature extraction may produce new features from original data.

4.1. Feature selection

It has been well recognized that the combinations of individually good features do not necessarily lead to good performance [78]. Feature selection methods make efforts to find the minimum subset of the input feature set, in order to enhance the efficiency of the network computing [79]. There are two basic frameworks to achieve feature selection: **wrapper** and **filter** [80], as shown in Fig. 7. The wrapper approach uses the performance of the learner as the evaluation criterion of feature set for feature selection. The filter approach uses evaluation criteria to enhance the correlation between features and classes, and to reduce the correlation between features. As the name implies, it's like a filter that screens out unnecessary features. The major difference between the two ways of feature selection mentioned above is whether to use learning algorithms. In general, the filter approach is of fast speed, while the wrapper approach has better accuracy.

According to the adopted methods, feature selection in wind energy forecasting can be divided into five categories: (1) correlation-based, (2) clustering-based, (3) information-based, (4) metaheuristic optimization-based and (5) other feature selection methods. In most cases, correlation, clustering and information-based feature selection are filter methods, while optimization algorithm-based are wrapper methods.

4.1.1. Correlation-based methods

The correlation-based feature selection methods consider the correlation between features or correlation between features and output.

The autocorrelation function (ACF) and partial correlation function (PACF) have been reported in digging the inherent extent of lag in datasets [59]. Liu et al. utilized ACF and PACF together with Granger causality test to first obtain the wind speed features which were highly correlated to the target data, and then selected the feature set as the input of predictor [81]. In their research, the relationship between temperature and wind speed was analyzed and results were quite interesting: the temperature would affect the wind speed lagged two and three (1 h and 1.5 h). Therefore, the input variables for predictor were selected as the historical wind speed 0.5 h and 1.5 h ahead, and the temperature 1.5 h ahead.

Grey system theory is also utilized to implement correlation analysis of features. Jiang et al. used Lag 1 to Lag 6 of two wind speed series collected from the adjacent wind turbine generators as original inputs and applied grey correlation analysis (GCA) to select useful fluctuation information [82]. They utilized ranking method to sort the grey relational degree of variables from large to small. The subset was selected from the first Nth variables. The ranking results showed historical wind speeds 10 min, 20 min, 30 min, and 40 min ahead collected from target location and 10 min ahead from its adjacent wind turbine were crucially representative.

Table 3
A summary of decomposition methods in wind energy forecasting.

Category	Subcategory	Characteristics or implementation details	Utilized algorithms in references
Decomposition	Wavelet-based	• Good localization properties in the time domain and frequency domain • Multiresolution analysis • Greatly depends on the structure of decomposition binary tree and the choice of wavelet function and decomposition level	WT [11], DWT [64], MODWT [33], WPD [65]
	EMD-based	• Subseries has an instantaneous frequency with a clear physical meaning • Traditional EMD has problems of endpoint effect and mode mixing • Adaptive, few hyperparameters • Many variants have been proposed to enhance its ability	EMD [66], EEMD [67], FEEMD [68], CEEMD [69], TVF-EMD [37], CEEMDAN [40], ICEEMDAN [70]
	Variational mode decomposition	• Multiple band-limited intrinsic mode functions • Non-recursive method	[71,72]
	Self-adaptive decomposition	• Divide the frequency band autonomously • Overcome the defect of predetermining parameters	OVMD [50], EWT [73]
	Hybrid decomposition	• Combination of different decomposition method • Extract more detailed components • Perform better in aspect of accuracy and stability • Not easy to make reasonable decomposition scheme • Increase of computation time	WT + SSA [58], WPD + VMD [62], WPD + EMD [55,63], WPD + FEEMD [54], WPD + CEEMDAN [74], EMD + WPD [56], CEEMD + VMD [75], CEEMDAN + VMD [57,76], VMD + SSA [77] , VMD + WPD [60], SSA + FEEMD + VMD [59]

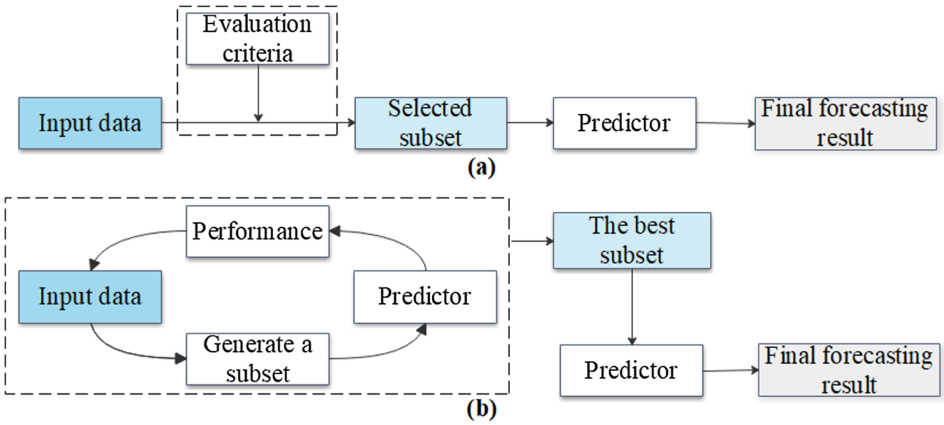

Fig. 7. The basic frameworks of feature selection: (a) filter, (b) wrapper.

In Ref. [64], a novel correlation-based feature selection model was subsumed within DWT decomposition. They first calculated the correlation coefficient of decomposed subseries and original data. If the correlation coefficient was smaller than the specified threshold, it would be discarded as illusive component. The original data was decomposed into nine subseries. The correlation coefficients showed that the 6th subseries had the highest absolute value (0.600). But the 4th subseries was only 0.004, which was lower than the selected threshold (0.06). Therefore, it was eliminated to achieve selection of feature sets.

4.1.2. Clustering-based methods

Some other feature selection methods are based on clustering techniques, which have been appealing increasing attention in recent years. The clustering-based feature selection methods first aggregate the input data into multiple clusters, then remove a few clusters from it, or select the clusters which are the most similar to the predicted target.

Yu et al. employed density-based spatial clustering of applications with noise (DBSCAN) to cluster the original training samples into several groups, which contain the outliers and others [83]. The utilized clustering method can effectively identify outliers that do not generally occur and are not representative, so that they can be eliminated to strengthen the forecasting ability. But the DBSCAN is sensitive to its two key parameters, i.e., ε and *MinPts*. In the literature, they chose to enlarge the *MinPts* value to 10, in order to sort out more outliers. The forecasting accuracy of the model with feature selection had a 30% improvement compared with models without feature selection.

Wang et al. used the k-means method to classify the dataset including wind speed, wind direction, temperature, humidity, and pressure into k clusters [84]. The clustering objects were vectors constructed by wind speed, temperature, humidity and pressure. They selected cluster samples which were similar to the target as the input of forecasting model. The proposed k-means feature selection method showed the greatest improvement in forecasting accuracy and efficiency of the proposed model. Azimi et al. proposed an improved k-means algorithm to cluster wind power data [85]. In the proposed algorithm, they used a new method to solve the problem of random initialization of cluster centroids. It showed better accuracy than traditional k-means clustering feature selection. Compared with the k-means approaches, spectral clustering (SC) is more adaptable to data and requires less computational load. Generated from graph theory, the SC converts the clustering problem into spectral partitioning in an undirected, weighted graph. The SC was employed to reduce the original 5840 samples to less than 1000, which greatly enhanced model efficiency[86].

In Ref. [87], kernel-based fuzzy c-means clustering (KFCM) was used to cluster the samples into similar fluctuation patterns. To ameliorate the problem of data discontinuity when intermittently using sample points, they adopted an improvement in the proposed clustering method by changing the clustering object from data points to data vectors. The comparison experiments showed that using the samples halved by KFCM to achieve forecasting can elicit better or similar performance. It implied that the KFCM feature selection is helpful in extracting the data traits.

4.1.3. Information-based methods

The reported information-based feature selection methods in wind energy forecasting include Mutual information (MI) [88] and conditional mutual information (CMI) [89,90], which are based on entropy. Fig. 8 shows the Venn diagram of relationships between correlated variables X and Y. MI estimates the level of information between two variables. The MI between X and Y is defined as follows:

$$I(X;\,Y) = \sum_{x\in X}\sum_{y\in Y} P(x,\,y)log\frac{P(x,\,y)}{P(x)P(y)} = H(X) - H(X|Y) \tag{1}$$

where $H(X)$ and $H(X|Y)$ are the entropy and conditional entropy, respectively.

CMI is an extension of MI. Given random variables X, Y and Z, the CMI is defined as:

$$I(X;\,Y|Z) = H(X|Z) - H(X|Y,\,Z) \tag{2}$$

The features selected by CMI have both independent information and weak mutual dependence. Huang compared the CMI with MI and Pearson Correlation Coefficient (PCC) [90]. The experimental redundancy feature set was the historical and statistical indicators of wind speed and meteorological factors on 21 January, 17 April, 18 July and 7 November 2009. The results proved that CMI could better enhance the performance of model than MI.

4.1.4. Metaheuristic optimization-based methods

Several research works have adopted metaheuristic optimization algorithms to achieve feature selection. The general concept is to regard the selection of subsets as a problem of random search optimization for

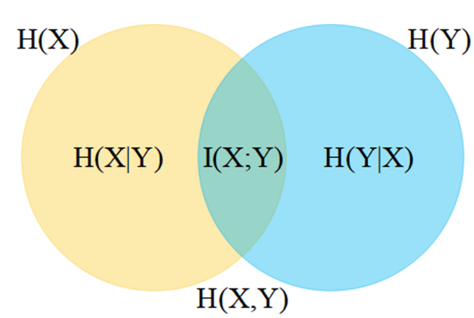

Fig. 8. The Venn diagram of relationships between correlated variables X and Y.

the original data set. The optimization process of input data is a typical wrapper feature selection. It uses metaheuristic optimization algorithms to generate a subset of input variables, and the subset is used to train a predictor, so as to obtain corresponding forecasting results. Then, the quality of the subset is evaluated the accuracy of forecasting result. This process loops multiple times and gets different subsets. Finally, the obtained subsets are compared to get the best subset. In general, the subset with the lowest forecasting error on validation set is selected as the best subset.

Salcedo-Sanz utilized coral reefs optimization algorithm (CRO), a bio-inspired approach to select the best set of input variables. Experiments of wind farms in the USA showed CRO-based model could a get better result [91]. The population-based optimization methods have been widely implemented for feature selection of wind speed series, such as particle swarm optimization (PSO) [92], genetic algorithm (GA) [93]. Ref. [94] compared two optimization algorithms, namely PSO and differential evolution (DE). The candidate inputs were time delay vectors of data collected from 30 wind farms and the selected input were used to forecast single wind farm in German. Compared with manually selecting input variables, the forecasting error showed a great reduction by using the proposed automated feature selection methods.

Some other optimization algorithms used for feature selection are introduced as binary versions to solve the optimization problem of discrete parameters. In Ref. [50], PACF was firstly used to determine lag order, and then the candidate input was selected using a binary-value backtracking search algorithm (BBSA). In Ref. [37], TVF-EMD decomposition was combined with feature selection. The decomposed IMFs were constructed to the candidate feature, and a novel hybrid population-based algorithm namely binary particle swarm optimization and gravitational search algorithm (BPSOGSA) was applied to achieve feature selection. The models with feature selection generally obtained lower forecasting error than models without feature selection.

4.1.5. Other feature selection methods

Besides the feature selection methods mentioned above, some other popular methods are also used. Their implementation details will be introduced in this subsection.

Phase space reconstruction (PSR) is a method based on the chaos theory, in which delay time is taken into consideration [95]. It is used to analyze the dynamic characteristics of time series and mapping the sequences into phase space. After the wind speed time series are transformed into the phase space, the data in phase space are used as input to carry out forecasting. Wang et al. used the C-C method to determine the embedding dimension and delay time of PSR and then utilized it to select the input form of wind speed series collected from Zhangye, Wuwei, Jiuquan, and Mazong Mountains in China [96]. Some scholars combined PSR with SE to achieve feature selection [68]. The original wind speed series was first decomposed into 8 IMFs and a residue by FEEMD, and then the subseries were recombined according to the SE values. The obtained five subsets were regarded as candidate input and transformed into data in phase space. A case study from two different wind farms in China verified the validity and effectiveness.

Random forest (RF) is a new and highly flexible machine learning algorithm, which has been gaining increasing attention. Its essence is actually the ensemble of numerous decision trees. In Ref. [32], an input feature selection method based on RF was adopted. The importance of candidate features including temperature, humidity, atmospheric pressure, and historical wind speed were evaluated by the RF. The proposed RF feature selection was based on the wrapper method. In every iteration, the forecasting error was calculated and after all iterations finished, the obtained feature subset had the fewest variables and the optimal forecasting results. It was concluded that the RF method effectively improved the forecasting ability of intelligent predictors by selecting the optimal input variables.

Gram–Schmidt orthogonal (GSO) can vectorize a set into a standard orthogonal vector group. In Ref. [47], GSO process was utilized to iteratively select feature which could maximize the cosine of the angle between the input feature vector and the output vector. It is noteworthy that they set the number of selected features by an optimization method called gravitational search algorithm (GSA), and the number of selected features with the lowest RMSE was adopted. Therefore, it's a combination of wrapper and filter methods.

However, most of the existing feature selection methods suffer from two major drawbacks: (a) linear methods like ACF and PACF only take linear relations into consideration; (b) nonlinear methods like KPCA take continuous feature subset as a factor without analyzing lag. To solve these problems, Feng et al. proposed a deep feature selection framework [97]. It combined four approaches: (a) PCA; (b) Granger causality test (GCT); (c) ACF and PACF; and (d) recursive feature elimination (RFE). They firstly conducted PCA to reduce the variable dimension and then the GCT was used to study the correlation between other variables and wind speed series. In order to explore the lag of variables, ACF and PACF were employed. At last, they utilized RFE to ensure effective integration with other variable time lags. The proposed deep feature selection framework is quite thought-provoking. As expected, the experiments using data from seven Surface Radiation (SURFRAD) locations proved that the proposed model could significantly improve 1-h ahead forecasting performance. Similarly, Wang et al. utilized Kullback-Leibler divergence (KLD), energy measure (EM) and SE to achieve optimal feature selection and extraction of decomposed wind power subseries [98]. They recombined the components into less subseries on the basis of SE value, which can improve calculation efficiency and forecasting accuracy. The classification of the feature selection methods discussed above is given in Table 4.

4.2. Feature extraction

4.2.1. Linear strategy

The most commonly used feature extraction method is principal component analysis (PCA). It is a linear feature extraction method. By

Table 4
A summary of feature selection methods in wind energy forecasting.

Category	Subcategory	Characteristics or implementation details	Utilized algorithms in references
Feature selection	Correlation-based	• Consider the correlation between features or correlation between features and output	Correlation-aided DWT [64], ACF and PACF [99], GSO [47], GCA [82]
	Clustering-based	• Aggregates the input data into multiple clusters • Strong interpretability	DBSCAN [83], KFCM [87], K-means [84,85], SC [86]
	Information-based	• Based on entropy • No need to make any assumption about the relationship between data	MI [100], CMI [90]
	Metaheuristic optimization-based	• Randomly search for optimal combinations from input variables • Good applicability • Not subject to specific data characteristics	BPSOGSA [37], BBSA [50], CRO [91], GA [93], PSO [94], DE [94]
	Others	• Some other popular feature selection methods, including hybridization of several algorithms	RF [32], PSR [101], Deep feature selection [97], KLD-EM-SE [98]

means of orthogonal transformations, the original random vector whose components are related to each other is transformed into a new random vector whose components are not mutually related. Then the dimensionality reduction is carried out on the multidimensional variable system so that it can be converted to the low-dimensional variable system with higher precision.

Many applications of PCA have been made in wind energy forecasting [102]. Kong et al. employed the PCA to extract the major variables actually affecting the wind speed [103]. Contribution rates of principal components were ranked as wind speed, temperature, air pressure, wind direction in descending order. If all principal components were concerned and chosen as input, the forecasting error indices were not the best, which articulated the redundancy of factors do no good to forecasting models. Furthermore, forecasting based on the combination of wind speed, temperature and air pressure had the best performance. Their experiment demonstrates that considering all factors affecting the wind speed is unnecessary and profitless. In Ref. [104], a 47-dimensional original vector was reduced into an18-dimensional space using generalized PCA. An interesting conclusion is that the PCA-based model outperformed the clustering-based model, demonstrating the effectiveness of feature extraction.

4.2.2. Nonlinear strategy

Kernel principal component analysis (KPCA) is a nonlinear version of PCA. It first carries on the nonlinear transformation to the sample, then adopts the PCA in the transformation space. KPCA generalizes PCA to nonlinear conditions through kernel function. Sun et al. developed an integrated approach for wind speed forecasting, combining PSR with KPCA [101]. The PSR was used to select the input vectors and the KPCA was used to reduce dimension. The experimental results showed that the proposed hybrid model significantly outperformed other benchmark models.

5. Data correction

This section reviews the data processing methods which are utilized to correct input data sets or forecasting results. Four totally different data correction methods have been adopted in the published papers, including denoising, residual error modeling, outlier detection and filter-based correction.

5.1. Denoising

Denoising of input data is an important way to improve the performance of forecasting models. The purpose of denoising is to sort out and strengthen the effective information of input data and remove the noise information that interferes with model training.

5.1.1. Singular spectrum analysis

Singular spectrum analysis (SSA) is a singular value decomposition (SVD) based method that can effectively decompose and reconstruct signals. It firstly converts the one-dimensional data into a trajectory matrix by an appropriate window length, and the major eigenvalues are selected to reconstruct components with significant tendency information. A detailed illustration of the crucial steps in SSA is given in Fig. 9. Apparently, the first several eigenvalues are representative of the statistically significant modes of wind speed. The reconstructed series reflects the variation tendency of wind energy, while the residual series mainly contains noise. The SSA has been widely applied to reduce the negative impact of random interference characteristics on raw data [77]. L. Xiao et al. compared SSA with FEEMD and concluded that the SSA-based model prevailed the FEEMD-based model in the aspect of accuracy, no matter one-step or multi-step forecasting [105]. In Ref. [106], in order to enhance the performance, an optimization method named brain storm optimization (BSO) was employed to determine the

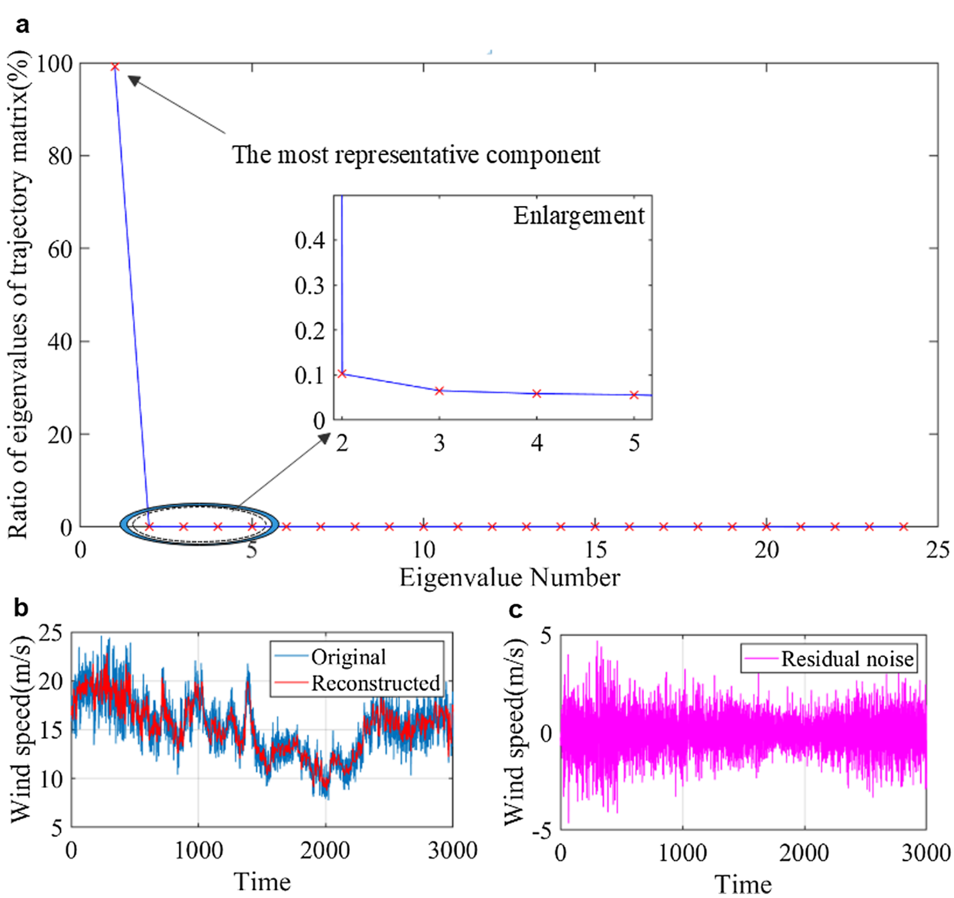

Fig. 9. The illustration of crucial steps in SSA. (a) Eigenvalues of trajectory matrix. The number of eigenvalues is equal to window length, which is set as 24 in this case. (b) Reconstructed series, determined by the selected eigenvalues. The red curve represents the trend component reconstructed by the first 5 eigenvalues. (c) Residual series reconstructed from the remaining 19 eigenvalues. It is usually considered as noise.

Table 5
The implementation details of decomposition-based denoising methods.

Decomposition method	Decomposition results	Discarded components	Reference
WT	An approximation component and a detailed component	The detailed component	[81]
WPD	Two approximation components and two detailed components	Highest-frequency component	[96]
EMD	Five IMFs and a residual	IMF1	[107]
CEEMD	Eight to ten IMFs and a residual	Highly volatile IMFs	[90]
CEEMDAN	Nine IMFs and a residual	high-frequency component	[108]
EWT	Three uncorrelated modes and a residual	The residual	[109]
EEMD	Seven IMFs and a residual	IMF1	[110]
	Seven IMFs and a residual	IMF1 and IMF2	[111]
	Five IMFs and a residual	IMF1 and residual	[112]

optimum window length and number of eigentriples of SSA. Experimental results showed that forecasting errors significantly decreased after BSO-SSA denoising.

5.1.2. Decomposition-based denoising

The decomposition-based denoising method is associated with Section 3. The original wind energy series is decomposed into several subseries and then some of the subseries are considered as noise and discarded. Table 5 lists the discarded components of decomposition-based denoising methods reported in the literature. Although the selected components are different, they all have one thing in common: they are components with the highest frequency in decomposition results. By discarding volatile components, the forecasting accuracy and model efficiency can be effectively improved [90].

The treatment of discarding IMF1 seems to sort of run counter to some hybrid decomposition algorithms mentioned in Section 3.5, which further decompose the IMF1 to achieve better forecasting precision. However, two ways of dealing with IMF1 all make sense. As is reported in Ref. [107], the high-frequency IMF1 is too small to influence the final forecasting result and its forecasting is unpredictable. Hence, some models choose to neglect the IMF1 component and regard it as noise, because the forecasting accuracy may be worse when IMF1 is taken into account (as found in Ref. [56]). While in hybrid decomposition-based models, the most disorder and irregular subseries IMF1 is further decomposed to decrease the forecasting difficulty, making it possible to predict.

5.1.3. Wavelet threshold denoising

Another denoising method used for wind speed series is wavelet threshold denoising (WTD), whose procedure is shown in Fig. 10. It firstly decomposes the original series into several subseries, and then the wavelet coefficients of each subseries are manipulated using an appropriate threshold. Since the wavelet coefficients of signals and noises have different intensity distributions at different scales, the wavelet coefficients of noises at each scale are removed and the wavelet coefficients of signals can be retained. Finally, the wavelet coefficients after processing are reconstructed by using the inverse wavelet transform to realize denoising. In Refs. [63,113], the WTD was used to denoise the original wind speed series and reduce the containing noise. The results showed that the it achieved more than 10% improvement of forecasting accuracy.

5.2. Residual error modeling

The residual error modeling is the most commonly used post-processing method in wind energy forecasting. Fig. 11 shows its general framework. It establishes a predictor to fit the forecasting residual error, and the trained residual error forecasting model is used in the test set. The forecasting error series is superimposed with the preliminary forecasting results to correct the systematic errors. Accordingly, the constructed models for original series and residual error are noted as the **principal predictor** and **subordinate predictor**, respectively. Residual error modeling is a process that uses a subordinate predictor to learn the changing trend of errors and correct the preliminary forecasting result obtained by the principal predictor.

Li et al. selected long short-term memory (LSTM) and regularized extreme learning machine (RELM) as principal and subordinate predictor, respectively [52]. The results showed that the residual error model could significantly improve forecasting performance. In Ref. [43], the residual series was decomposed into several IMFs by ICEEMDAN. They built an ARIMA model for each IMF. The proposed residual error modeling is a common-used structure for decomposition-based wind speed forecasting.

Some researchers used models with different characteristics to model linear and nonlinear components of wind speed series, respectively. The hypothesis that there is an additive relationship between the

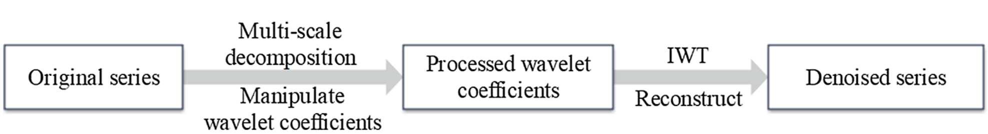

Fig. 10. The procedure of WTD.

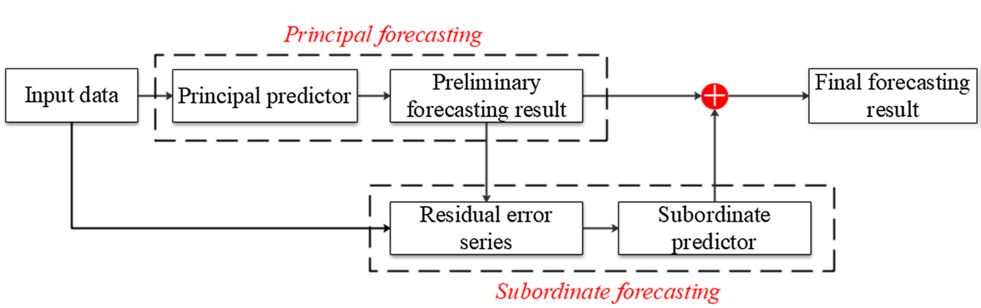

Fig. 11. The general framework of residual error modeling.

Table 6
The results of published papers using residual error modeling.

Reference	Principal predictor	Subordinate predictor	Data sources	Forecasting horizon	Improvement percentage[a]
[98]	LSTM	LSTM & GARCH	Hexi corridor of China	1–5 h	24.56–10.75% (MAE)
[116]	LSSVM	Markov	Northwest China	1 d	2.67% (MAPE)
[28]	ARIMA	ANN/SVM	Colorado, USA	1–9 h	4.99% to −0.18%
[43]	ELM	ICEEMDAN-ARIMA	Hexi Corridor of China	10 min	29.75% (MAE)
[52]	LSTM	RELM	Xinjiang, China	10–50 min	72.18–63.83% (MAPE)
[64]	LSSVM	GARCH	Minnesota, USA	30 min	1.82% (RMSE)
[114]	ARMA	BP/SVM/RF	Six observation stations in China	1 h	23.6% (MAE)
[115]	SVM	ELM/SVM	Hebei Province, China	1–6 h	23.94–36.74% (NMAE)

[a] The improvement percentage is based on the error indices of forecasting using residual error modeling E_1 and forecasting without residual error modeling E_2. $Improvement\ percentage = \frac{E_2 - E_1}{E_2} \times 100\%$.

linear and nonlinear component of wind speed series is assumed [28], which can be expressed as follows:

$$y_t = \widehat{L}_t + \widehat{N}_t \tag{3}$$

where y_t is the final forecasting result, $\widehat{L}_t$ is the linear component and $\widehat{N}_t$ is the nonlinear component.

Shi et al. used the ARIMA model to capture the linear component of wind speed series. The residual series were fitted by nonlinear models including ANN and SVM [28]. The combination of a non-parametric (NP) model and three AI/ML models were applied [114]. In the proposed hybrid-ARMA-NP model, ARMA is used to obtain the linear forecasting result, and NP is used to capture the nonlinear components. AI/ML models including BP, SVM and RF are used for comparison. In the proposed hybrid-NP-ARMA model, NP and AI/ML models are used as the principal model, while ARMA is used as the subordinate model.

The proposed residual error modeling methods mentioned above simply forecast the residual series. Nevertheless, some decision methods can be made before residual error modeling. Correlation analysis has been implemented, and the SVM or ELM model was utilized to model the residual according to the results of correlation analysis [115]. In Ref. [98], the correlation and heteroscedasticity of residual error were considered. They divide the states of error components into four kinds: (a) residual series has both correlation and heteroscedasticity, (b) residual series only has correlation, (c) residual series only has heteroscedasticity, and (d) correlation and heteroscedasticity are all not obvious in residual series. For correlation, they applied the LSTM to forecast the residual error, while for heteroscedasticity they used generalized autoregressive conditional heteroskedasticity (GARCH). The models with residual error modeling outperformed other models for one-step, three-step and five-step ahead forecasting.

Markov model has also been employed to correct the forecasting error in Ref. [116], which employs state transition probability to calculate the state of forecasting error. They found that the forecasting results after error modeling were sometimes poorer than that of before error modeling. Therefore, the mean values of original and corrected forecasting result were taken as the final forecasting result. The average strategy was proved to be more powerful than other models. Table 6 lists the result of published papers using residual error modeling methods. In most of the proposed models, residual error modeling can improve the accuracy by more than 10 percent, which implies its effectiveness.

5.3. Outlier detection

In general, outliers of time series are data points that do not follow a general pattern of change or historical trend. When the data is contaminated by outliers, the data-driven models may mistakenly learn the abnormal information of the data itself, and it is difficult to achieve high accuracy in the prediction based on the data set.

There are two main solutions utilized to reduce the negative impact of outliers on the model, including outlier detection methods and robust regression models [117]. The outlier detection methods serve as a preprocessing method to detect and correct the outliers. While the robust regression models don't detect or treat outliers directly, but control the influence of anomaly through loss function or error distributions. More details of robust regression models for wind speed forecasting can be found in Refs. [117–120]. Here we focus on outlier detection preprocessing methods. The related publications for wind energy forecasting are relatively rare. However, the general steps of outlier detection processing methods are consistent. They amend specific points in forecasting series according to certain rules. The implementation process includes two steps: (a) detection and (b) correction or removal.

5.3.1. Local approaches

The local approaches use sliding window or local density to detect the anomalies of data points. These kinds of outlier detection methods are sensitive to the parameters of models. In Ref. [121], the collected wind data were divided into six categories, i.e., valid data, missing data, constant data, exceeding data, irrational data and unnatural. The irrational data meant illogical in physics and the unnatural data represented low wind power output during periods of high wind speed. These two categories deviated greatly from the majority patterns and were considered as outliers. The local outlier factor (LOF) algorithm and similarity measurement were used to detect and remove outliers, which identify whether the point is abnormal by comparing the density of each point and its neighborhood points. If the density of a point is lower, it is more likely to be identified as an abnormal point. Numerical experiments based on the data collected from northwest China verified the effectiveness of the proposed outlier detection method.

Outlier detection has also been used in postprocessing. The train of thought is to alleviate the over-fitting phenomenon of the predictor, based on actual observations of wind speed series. Mi et al. proposed a time-efficient outlier detection method based on a sliding window [63]. The upper and lower limits were determined with m observed values to decide whether the predicted value was an outlier. If the predicted value was an outlier, it would be corrected as the mean value of n observations. The proposed outlier detection postprocessing method significantly prevailed other models especially when the forecasting horizons were expanded to five steps ahead.

5.3.2. Regression-based approaches

Wang et al. proposed two regression-based methods for outlier detection and correction of wind speed series [122,123]. In [122], they adopted the weighted least squares algorithm (WLS) to determine the outliers. If the studentized residual of regression $t_i > 1.96$, the observed value at the ith point was seen as an outlier and it would be corrected by cubic spline interpolation algorithm. Another proposed method is based on support vector regression (SVR) [123]. They firstly constructed the SVR model to predict the original wind speed series and then calculated the residual series. If the sample residuals $E_i > \sigma$, the ith

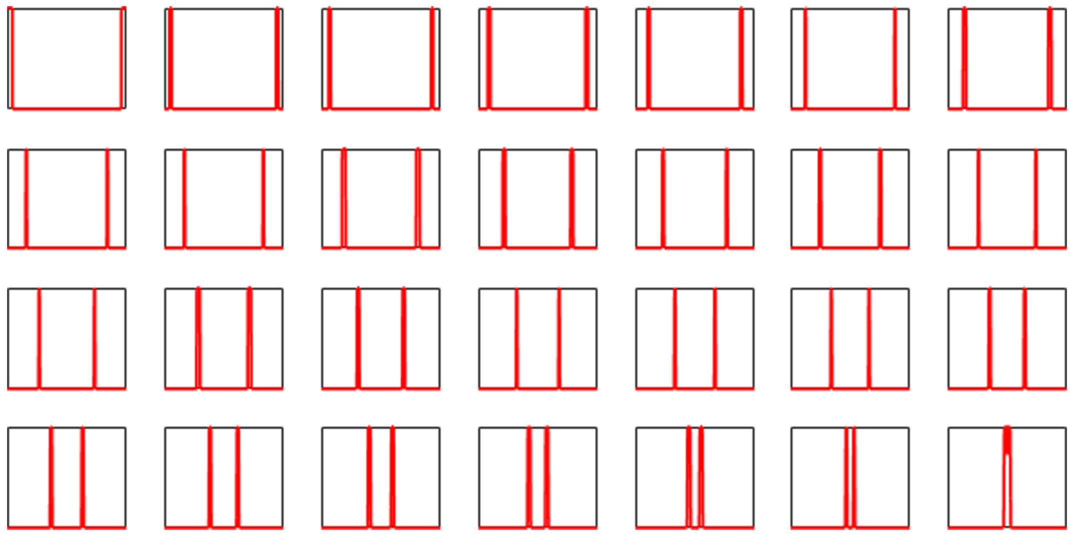

Fig. 12. An empirical filter bank of EWT in the Fourier domain. In this case, 28 band-pass filters are obtained.

sample was considered an outlier and replaced by the forecasted value. Both of the proposed models were verified to be effective for daily wind speed forecasting.

5.4. Filter-based correction

Filtering is an operation to filter the frequency of a specific band in the signal. It is an important measure to suppress and prevent interference. The basic assumption of filter-based correction methods is that the frequency band of forecasting series is consistent with that of the original series. The spectrum segment of the original series is taken as the standard to filter error, so as to remove the extra components generated by forecasting.

Wavelet filter (WF) has been utilized to correct wind speed series forecasting results [62]. The authors removed the excrescent frequency component using appropriate thresholds to dispose wavelet coefficients, which showed great performance in multi-step forecasting results. Besides, rather than just sum up forecasting subseries as most of the decomposition-based models usually do [31], some inverse methods have been successfully applied, mainly including inverse WT (IWT) [88] and inverse EWT (IEWT) [52,73]. The inverse transforms are performed according to the filter bank, as shown in Fig.12. The forecasting results of subseries are transformed back into the original domain, so to get the reconstructed series. They are regarded as methods to filter out abnormal points in forecasting series, which have been reported to avoid the unexpected forecasting values and improve model stability, as well as forecasting accuracy [73].

6. Discussion

A general analysis of data processing methods in wind energy forecasting has been presented in this paper. They are clearly classified as decomposition, feature selection, feature extraction, denoising, outlier detection and filter-based correction. Since the attention and application of each method are different, the usage frequencies are also different, as shown in Fig. 13. Apparently, decomposition is the most commonly used data processing method. This is because its effect on the forecasting accuracy is the most obvious. Some models using decomposition methods can reduce the forecasting error by more than 50% when compared with models without decomposition algorithms [46,50]. Nevertheless, filter-based correction does not gain much attention as other methods do. However, it was proved to bring more than 30% improvement in forecasting accuracy [52]. Its potential in wind energy forecasting, especially in decomposition-based models can be further explored. In the seven categories of data processing methods, the literature of outlier detection is also not relatively extensive. This is mainly because of the difficulty in identifying whether a sample is an outlier or a normal point. Mistaking the normal data as an outlier and performing the correction process may destroy the inherent regularity of the wind energy series, resulting in even worse forecasting accuracy. To address this concern, more robust identifiers for outliers can be developed and experimented in the future.

Table 7 gives a general evaluation of these methods from six different aspects, including accuracy improvement, usage frequency, consuming time, robustness to parameters, maturity and implementation difficulty. It is noteworthy that because the data processing methods are not mutually exclusive or antagonistic, no one method is absolutely superior to the others. In fact, they are sometimes combined together to construct more powerful forecasting models [64].

6.1. Possible development trends

In the future, increasing data-driven models will be proposed for wind energy forecasting and the data processing methods need more attention because the historical data are always the fundamental of predictors. There are some viewpoints that can be future research directions of data processing methods in wind energy forecasting. They are listed as follows:

6.1.1. Exogenous input variables

Most of the published papers only adopted the historical wind speed or power series to complete the forecasting tasks. They can be referred to as

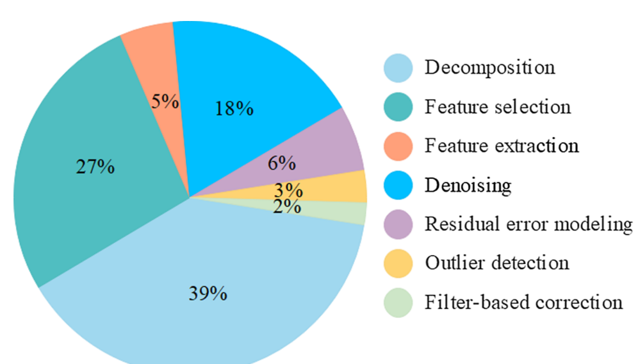

Fig. 13. The usage frequencies of data processing in wind energy forecasting models: results derived from 130 application cases.

Table 7
The performance evaluation of data processing methods in wind energy forecasting.

Category	Accuracy improvement	Usage frequency	Consuming time	Robustness to parameters	Maturity	Implementation difficulty
Decomposition	****	*****	*****	**	****	**
Feature selection	***	****	***	**	***	**
Feature extraction	**	**	**	***	***	**
Denoising	*	***	*	****	****	*
Residual error modeling	**	**	***	***	**	***
Outlier detection	***	*	*	*	*	***
Filter-based correction	***	*	*	****	***	*

models without exogenous variables. However, subsuming some exogenous variables into the forecasting models may be able to enhance the interpretability. Meteorological conditions are important factors that can cause wind. Some researchers have virtually taken this into consideration and selected different exogenous variables as the input of training models.

However, considering the exogenous variables may sometimes induce the curse of dimensionality. Therefore, some dimensionality reduction methods are also adopted. The selected input variables and dimensionality reduction methods in wind energy forecasting models are listed in Table 8 for reference. With the continuous development of deep learning, more data and feature sets about wind energy are needed to improve the forecasting accuracy further.

6.1.2. Applications of distributed computing

With the expansion of data sets, parallel computing of large-scale data has gradually gained attention. Reasonable and efficient distributed parallel algorithms have become an important method to effectively utilize the processor in the distributed environment and improve the performance of the distributed system. Recently, the Apache Spark framework was employed to forecast big data time series, which demonstrated great viability and scalability [127]. However, it has never been used in the field of wind energy forecasting. More related parallel algorithms should be investigated. The Apache Spark is a possible application for the distributed computing of wind energy forecasting. It is a fast computing engine specially designed for large-scale data processing. It takes full advantage of clustering for high-speed computing and storage. By running programs distributedly in clusters, the computational efficiency can be greatly enhanced. The Apache Spark is easy to be integrated with other open source products, such as the Apache Mesos which can be directly used as resource management. Besides, the data can be accessed from the data storage system without data migration, such as the Hadoop Distributed File System (HDFS). The applications of distributed computing also fit the future development trend of big data.

6.1.3. Applications of deep learning-based feature extraction

Deep learning is a powerful method and it has been successfully applied in various fields. It stacks multiple layers and utilizes a deeper structure to enhance the learning ability of models. In wind energy forecasting, deep learning has been used to construct predictors in several research papers [85]. However, deep learning-based feature extraction has never been developed to process wind data, although some applications can be found in other fields. For instance, in Ref. [128], an autoencoder (AE) was utilized to realize the feature extraction of solar power data. In their proposed model, the outputs of encoder were considered as the features of solar power, and another deep recurrent neural network, i.e., LSTM was adopted to model the encoding part of AE. It showed great superiority in extracting features from solar power.

Compared with traditional feature extraction approaches, deep learning techniques can extract the hidden representation of data due to its deeper structure and better capability. This poses great significance to the applications of deep learning-based feature extraction, especially for the data-driven models which require massive useful features of wind.

6.2. Challenges in the application of data processing methods

Although various data processing methods have been extensively employed in recent years, there are still some challenges that exist in the current applications. They need to be noticed and further studied.

6.2.1. Adaptively evaluate the postprocessing results

In the postprocessing methods like residual error modeling and outlier detection, the overcorrection of data is possible to happen. However, the current applications rarely concern about the rational evaluation of processing results. In this way, the accuracy on testing data may be worse, when an overcorrection of preliminary forecasting result occurs. Some evaluation criteria or methods for data correction

Table 8
The wind energy forecasting models with exogenous input variables.

Reference	Variables	Dimensionality reduction
Without dimensionality reduction		
Gallego [124]	Wind power, wind speed, wind direction	–
Camelo [125]	Wind speed, pressure, temperature, precipitation	–
De Giorgi [126]	Wind power, pressure, temperature, relative humidity	–
Lydia [23]	Wind speed, wind direction, solar radiation, temperature	–
With dimensionality reduction		
Azimi [85]	Wind power, wind direction, air temperature	Feature selection: T.S.B K-means
Wang [32]	Wind speed, temperature, humidity, atmospheric pressure	Feature selection: RF
Cong [97]	Wind speed, temperature, humidity, wind direction, pressure	Deep feature selection
Wang [84]	Wind power, wind direction, temperature, humidity, pressure	Feature selection: K-means
Huang [90]	Wind speed, temperature, relative humidity, absolute humidity, atmospheric pressure, wind shear	Feature selection: CMI
Liu [86]	Wind speed, temperature, characters (maximum, minimum, mean)	Feature extraction: PCA Feature selection: SC
Sun [101]	Wind speed, air temperature, air pressure, relative humidity	Feature selection: PSR Feature extraction: KPCA
Salcedo-Sanz [91]	Wind direction at 5 different heights, temperature at 3 different heights, humidity, sea level pressure, long wave down radiation, short wave down radiation, precipitation, 8 function combinations of variables	Feature selection: CRO

results can be developed to ensure the effectiveness of postprocessing algorithms.

6.2.2. Choose diverse predictors to forecast different subseries

The decomposition-based hybrid models generally adopt the same predictor to forecast all subseries. They neglect the diversity of subseries with different frequencies. How to adaptively choose the optimal predictor for each subseries is a problem worthy of study. Taking the characteristics of subseries into consideration, the corresponding forecasting results can be more accurate. From the perspective of model interpretation, this can also be a breakthrough in improving performance.

6.2.3. Fully utilize the original high-resolution dataset

The original dataset of the existing data-driven models is generally high-resolution series with a second-level sample rate, such as 1 s [124], 3 s [62] and 5 s [14]. To remove the redundant data and achieve longer forecasting horizons, a simple method is usually employed to convert the original data into low-resolution results, that is, average. It can be regarded as a filter method that is able to remove fast fluctuations [117]. Nevertheless, there is one notable fact that when processing the high-resolution data by average, extensive useful information is lost as well. The approaches to fully explore and utilize the original high-resolution dataset should be analyzed.

7. Conclusions

In the face of increasingly serious environmental problems, wind speed and wind power forecasting are vital for the management and utilization of wind energy resources. The existing data-driven forecasting models attach great significance to the proper application of data processing methods. In this paper, seven types of data processing methods reported in the literature were summarized, including decomposition, feature selection, feature extraction, denoising, residual error modeling, outlier detection and filter-based correction. The methodologies, functions, performance of each kind of data processing method were thoroughly investigated. They were evaluated from different aspects including accuracy improvement, consuming time and maturity. Moreover, some possible development trends, prospects and challenges were also presented. Several major conclusions can be extracted from the present review:

- In the classified seven categories, outlier detection and filter-based correction are relatively less used, accounting for only 3% and 2% of the employed methods, respectively. More attention can be paid to the proper designs of these postprocessing methods. Besides, to ensure their effectiveness, the models should try to avoid the overcorrection of data.
- There is no absolutely superior method, and it is impractical to develop only a data processing strategy to deal with various prediction scenarios. For instance, decomposition-based models often require more computational time because of the specific framework. Although they have been proved to be exceedingly effective in improving the prediction accuracy, the time efficiency of the model is sacrificed. Therefore, they are not well suited for predictive tasks with high real-time requirements.
- Different data processing methods are not mutually antagonistic. Some hybrid models subsume data decomposition and feature selection at the same time. With the development of different types of methods, more possible combinations of them may be proposed to further improve the model performance.

With the advent of the big data era, ongoing investigations of data processing technologies are required to better explore the potential of data-driven hybrid models, and lay a good foundation for further improvement of forecasting precision and efficiency.

Acknowledgements

This study is fully supported by the National Natural Science Foundation of China (Grant No. 61873283), the Changsha Science & Technology Project and Training Program for Excellent Young Innovators of Changsha (Grant No. KQ1707017), the Shenghua Yu-ying Talents Program of the Central South University and the innovation driven project of the Central South University (Project No. 2019CX005).

References

[1] Wang J, Shahidehpour M, Li Z. Security-constrained unit commitment with volatile wind power generation. Ieee T Power Syst 2008;23(3):1319–27.
[2] Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. Renew Sustain Energy Rev 2014;31:762–77.
[3] Lawan S, Abidin W, Chai W, Baharun A, Masri T. Different models of wind speed prediction; a comprehensive review. Int J Sci Eng Res 2014;5(1):1760–8.
[4] Cheng WY, Liu Y, Bourgeois AJ, Wu Y, Haupt SE. Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation. Renewable Energy 2017;107:340–51.
[5] Pinson P. Wind energy: Forecasting challenges for its operational management. Stat Sci 2013;28(4):564–85.
[6] Liu H, Tian H-Q, Li Y-F. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl Energy 2012;98:415–24.
[7] Akcay H, Filik T. Short-term wind speed forecasting by spectral analysis from long-term observations with missing values. Appl Energy 2017;191:653–62.
[8] Wang Y, Xie Z, Hu Q, Xiong S. Correlation aware multi-step ahead wind speed forecasting with heteroscedastic multi-kernel learning. Energy Convers Manage 2018;163:384–406.
[9] Hui L, 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.
[10] Xiao L, Dong YX, Dong Y. An improved combination approach based on Adaboost algorithm for wind speed time series forecasting. Energy Convers Manage 2018;160:273–88.
[11] Yu C, Li Y, Bao Y, Tang H, Zhai G. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Convers Manage 2018;178:137–45.
[12] Wang H-Z, Li G-Q, Wang G-B, Peng J-C, Jiang H, Liu Y-T. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[13] Chitsazan MA, Fadali MS, Trzynadlowski AM. Wind speed and wind direction forecasting using echo state network with nonlinear functions. Renew Energy 2019;131:879–89.
[14] Hu Q, Zhang R, Zhou Y. Transfer learning for short-term wind speed prediction with deep neural networks. Renew Energy 2016;85:83–95.
[15] Cheng WY, Liu Y, Liu Y, Zhang Y, Mahoney WP, Warner TT. The impact of model physics on numerical wind forecasts. Renew Energy 2013;55:347–56.
[16] Hu YL, Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm. Energy Convers Manage 2018;173:123–42.
[17] 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.
[18] Li C, Lin S, Xu F, Liu D, Liu J. Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in Northwest China. J Cleaner Prod 2018;205:909–22.
[19] 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.
[20] Jiang P, Yang H, Heng J. A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting. Appl Energy 2019;235:786–801.
[21] Hao Y, Tian C. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. Appl Energy 2019;238:368–83.
[22] Liu H, Duan Z, Li Y, Lu H. A novel ensemble model of different mother wavelets for wind speed multi-step forecasting. Appl Energy 2018;228:1783–800.
[23] Lydia M, Kumar SS, Selvakumar AI, Kumar GEP. Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers Manage 2016;112:115–24.
[24] Yan J, Liu Y, Han S, Wang Y, Feng S. Reviews on uncertainty analysis of wind power forecasting. Renew Sustain Energy Rev 2015;52:1322–30.
[25] Okumus I, Dinler A. Current status of wind energy forecasting and a hybrid method for hourly predictions. Energy Convers Manage 2016;123:362–71.
[26] Marugán AP, Márquez FPG, Perez JMP, Ruiz-Hernández D. A survey of artificial neural network in wind energy systems. Appl Energy 2018;228:1822–36.
[27] Wang J, Song Y, Liu F, Hou R. Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. Renew Sustain Energy Rev 2016;60:960–81.
[28] Shi J, Guo J, Zheng S. Evaluation of hybrid forecasting approaches for wind speed

and power generation time series. Renew Sustain Energy Rev 2012;16(5):3471–80.
[29] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34(6):243–54.
[30] Zendehboudi A, Baseer M, Saidur R. Application of support vector machine models for forecasting solar and wind energy resources: A review. J Cleaner Prod 2018;199:272–85.
[31] Qian Z, Pei Y, Zareipour H, Chen N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Appl Energy 2019;235:939–53.
[32] Wang H, Sun JX, Sun JB, Wang JL. Using random forests to select optimal input variables for short-term wind speed forecasting models. Energies 2017;10(10):1522.
[33] Li FD, Liao HY. An intelligent method for wind power forecasting based on integrated power slope events prediction and wind speed forecasting. Ieej T Electr Electr 2018;13(8):1099–105.
[34] Sun SZ, Fu JQ, Zhu F. A short term wind speed forecasting method using signal decomposition and extreme learning machine. Comm Com Inf Sc 2017;763:22–31.
[35] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc Roy Soc London Ser A: Math, Phys Eng Sci 1998;454(1971):903–95.
[36] Wang J, Zhang W, Li Y, Wang J, Dang Z. Forecasting wind speed using empirical mode decomposition and Elman neural network. Appl Soft Comput 2014;23:452–9.
[37] Zheng WQ, Peng XG, Lu D, Zhang D, Liu Y, Lin ZH, et al. Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: A new approach. Energy Convers Manage 2017;151:737–52.
[38] Huang YS, Liu SJ, Yang L. Wind speed forecasting method using EEMD and the combination forecasting method based on GPR and LSTM. Sustainab-Basel 2018;10(10):3693.
[39] Yang ZS, Wang J. A combination forecasting approach applied in multistep wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm. Appl Energy 2018;230:1108–25.
[40] Zhang WY, Qu ZX, Zhang KQ, Mao WQ, Ma YN, Fan X. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Convers Manage 2017;136:439–51.
[41] Li X, Li C. Improved CEEMDAN and PSO-SVR modeling for near-infrared non-invasive glucose detection. Comput Math Methods Med 2016;2016.
[42] Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomed Signal Process Control 2014;14:19–29.
[43] Wang L, Li X, Bai Y. Short-term wind speed prediction using an extreme learning machine model with error correction. Energy Convers Manage 2018;162:239–50.
[44] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62(3):531–44.
[45] Zhou JZ, Liu H, Xu YH, Jiang W. A hybrid framework for short term multi-step wind speed forecasting based on variational model decomposition and convolutional neural network. Energies 2018;11(9):2292.
[46] Zhou JZ, Sun N, Jia BJ, Peng T. A novel decomposition-optimization model for short-term wind speed forecasting. Energies 2018;11(7):1752.
[47] Li CS, Xiao ZG, Xia X, Zou W, Zhang C. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. Appl Energy 2018;215:131–44.
[48] Wang DY, Luo HY, Grunder O, Lin YB. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. Renew Energy 2017;113:1345–58.
[49] Wang XD, Yu QB, Yang Y. Short-term wind speed forecasting using variational mode decomposition and support vector regression. J Intell Fuzzy Syst 2018;34(6):3811–20.
[50] Zhang C, Zhou JZ, Li CS, Fu WL, Peng T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. Energy Convers Manage 2017;143:360–76.
[51] Gilles J. Empirical wavelet transform. IEEE Trans Signal Process 2013;61(16):3999–4010.
[52] Li YF, Wu HP, Liu H. Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction. Energy Convers Manage 2018;167:203–19.
[53] Liu H, Mi XW, Li YF. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manage 2018;156:498–514.
[54] Liu H, Tian H-Q, Liang X-F, Li Y-F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Appl Energy 2015;157:183–94.
[55] Liu H, Mi X, Li Y. An experimental investigation of three new hybrid wind speed forecasting models using multi-decomposing strategy and ELM algorithm. Renew Energy 2018;123:694–705.
[56] Yin H, Dong Z, Chen Y, Ge J, Lai LL, Vaccaro A, et al. An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. Energy Convers Manage 2017;150:108–21.
[57] Peng T, Zhou JZ, Zhang C, Zheng Y. Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine. Energy Convers Manage 2017;153:589–602.
[58] Yu CJ, Li YL, Zhang MJ. An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on Elman neural network. Energy Convers Manage 2017;148:895–904.
[59] Wang J, Wang Y, Li Y. 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.
[60] Zhang D, Peng X, Pan K, Liu Y. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine. Energy Convers Manage 2019;180:338–57.
[61] Tian C, Hao Y, Hu J. A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization. Appl Energy 2018;231:301–19.
[62] Liu H, Duan Z, Han FZ, Li YF. Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. Energy Convers Manage 2018;156:525–41.
[63] Mi XW, Liu H, Li YF. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. Energy Convers Manage 2017;151:709–22.
[64] Jiang Y, Huang G, Peng X, Li Y, Yang Q. A novel wind speed prediction method: Hybrid of correlation-aided DWT, LSSVM and GARCH. J Wind Eng Ind Aerod 2018;174:28–38.
[65] Liu H, Mi X, Li Y. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. Energy Convers Manage 2018;166:120–31.
[66] Niu DX, Liang Y, Hong WC. Wind speed forecasting based on EMD and GRNN optimized by FOA. Energies 2017;10(12):2001.
[67] Bokde N, Feijoo A, Kulat K. Analysis of differencing and decomposition preprocessing methods for wind speed prediction. Appl Soft Comput 2018;71:926–38.
[68] Sun W, Wang YW. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. Energy Convers Manage 2018;157:1–12.
[69] Du P, Wang JZ, Guo ZH, Yang WD. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. Energy Convers Manage 2017;150:90–107.
[70] Du P, Wang J, Yang W, Niu T. A novel hybrid model for short-term wind power forecasting. Appl Soft Comput 2019;80:93–106.
[71] Yang W, Wang J, Lu H, Niu T, Du P. Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: a case study in China. J Cleaner Prod 2019;222:942–59.
[72] Li R, Jin Y. A wind speed interval prediction system based on multi-objective optimization for machine learning method. Appl Energy 2018;228:2207–20.
[73] Liu H, Wu HP, Li YF. Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction. Energy Convers Manage 2018;161:266–83.
[74] Liu H, Mi XW, Li YF. Comparison of two new intelligent wind speed forecasting approaches based on Wavelet Packet Decomposition, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Artificial Neural Networks. Energy Convers Manage 2018;155:188–200.
[75] Wang JZ, Yang WD, Du P, Niu T. A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. Energy Convers Manage 2018;163:134–50.
[76] Yu CJ, Li YL, Zhang MJ. Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting. Energy Convers Manage 2017;147:75–85.
[77] Liu H, Mi XW, Li YF. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manage 2018;159:54–64.
[78] Cover TM. The best two independent measurements are not the two best. IEEE Trans Syst, Man, Cybernet 1974;1:116–7.
[79] Abedinia O, Amjady N, Zareipour H. A new feature selection technique for load and price forecast of electrical power systems. Ieee T Power Syst 2017;32(1):62–74.
[80] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27(8):1226–38.
[81] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2014;62:592–7.
[82] Jiang P, Wang Y, Wang JZ. Short-term wind speed forecasting using a hybrid model. Energy 2017;119:561–77.
[83] Yu CJ, Li YL, Xiang HY, Zhang MJ. Data mining-assisted short-term wind speed forecasting by wavelet packet decomposition and Elman neural network. J Wind Eng Ind Aerod 2018;175:136–43.
[84] Wang K, Qi X, Liu H, Song J. Deep belief network based k-means cluster approach for short-term wind power forecasting. Energy 2018;165:840–52.
[85] Azimi R, Ghofrani M, Ghayekhloo M. A hybrid wind power forecasting model based on data mining and wavelets analysis. Energy Convers Manage 2016;127:208–25.
[86] Liu D, Wang J, Wang H. Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. Renew Energy 2015;78:599–608.
[87] He QQ, Wang JZ, Lu HY. A hybrid system for short-term wind speed forecasting. Appl Energy 2018;226:756–71.
[88] Osório G, Matias J, Catalão J. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. Renew Energy 2015;75:301–7.
[89] Li S, Wang P, Goel L. Wind power forecasting using neural network ensembles with

feature selection. IEEE Trans Sustain Energy 2015;6(4):1447–56.
[90] Huang NT, Xing EK, Cai GW, Yu ZY, Qi B, Lin L. Short-term wind speed forecasting based on low redundancy feature selection. Energies 2018;11(7):1638.
[91] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, Blanco-Aguilera A, García-Herrera R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–Extreme learning machine approach. Energy Convers Manage 2014;87:10–8.
[92] Jursa R. Variable selection for wind power prediction using particle swarm optimization. Proceedings of the 9th annual conference on Genetic and evolutionary computation. 2017. p. 2059–65.
[93] Gupta R, Kumar R, Bansal AK. Selection of input variables for the prediction of wind speed in wind farms based on genetic algorithm. Wind Eng 2011;35(6):649–60.
[94] Jursa R, Rohrig K. Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. Int J Forecast 2008;24(4):694–709.
[95] Takens F. Detecting strange attractors in turbulence. Dynamical systems and turbulence, Warwick 1980. Springer; 1981. p. 366–81.
[96] Wang J-Z, Wang Y, Jiang P. The study and application of a novel hybrid forecasting model–A case study of wind speed forecasting in China. Appl Energy 2015;143:472–88.
[97] Cong F, Cui M, Hodge BM, Jie Z. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 2017;190:1245–57.
[98] Wang J, Li Y. Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl Energy 2018;230:429–43.
[99] Hu J, Heng J, Tang J, Guo M. Research and application of a hybrid model based on Meta learning strategy for wind power deterministic and probabilistic forecasting. Energy Convers Manage 2018;173:197–209.
[100] Liu T, Wei H, Zhang K. Wind power prediction with missing data using Gaussian process regression and multiple imputation. Appl Soft Comput 2018;71:905–16.
[101] Sun SL, Qiao H, Wei YJ, Wang SY. A new dynamic integrated approach for wind speed forecasting. Appl Energy 2017;197:151–62.
[102] Davò F, Alessandrini S, Sperati S, Delle Monache L, Airoldi D, Vespucci MT. Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. Sol Energy 2016;134:327–38.
[103] Kong X, Liu X, Shi R, Lee KY. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[104] Hu Q, Su P, Yu D, Liu J. Pattern-based wind speed prediction based on generalized principal component analysis. Ieee T Sustain Energ 2014;5(3):866–74.
[105] Xiao LY, Qian F, Shao W. Multi-step wind speed forecasting based on a hybrid forecasting architecture and an improved bat algorithm. Energy Convers Manage 2017;143:410–30.
[106] Ma XJ, Jin Y, Dong QL. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. Appl Soft Comput 2017;54:296–312.
[107] Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012;37(1):241–9.
[108] Niu X, Wang J. A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. Appl Energy 2019;241:519–39.
[109] Hu J, Wang J, Ma K. A hybrid technique for short-term wind speed prediction. Energy 2015;81:563–74.
[110] Santhosh M, Venkaiah C, Kumar DV. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. Energy Convers Manage 2018;168:482–93.
[111] Zhang KQ, Qu ZX, Wang JZ, Zhang WY, Yang FY. A novel hybrid approach based on cuckoo search optimization algorithm for short-term wind speed forecasting. Environ Prog Sustain 2017;36(3):943–52.
[112] Hu J, Wang J, Zeng G. A hybrid forecasting approach applied to wind speed time series. Renew Energy 2013;60:185–94.
[113] Cheng L, Zang H, Ding T, Sun R, Wang M, Wei Z, et al. Ensemble recurrent neural network based probabilistic wind speed forecasting approach. Energies 2018;11(8):1958.
[114] Han QK, Meng FM, Hu T, Chu FL. Non-parametric hybrid models for wind speed forecasting. Energy Convers Manage 2017;148:554–68.
[115] Liang Z, Liang J, Wang C, Dong X, Miao X. Short-term wind power combined forecasting based on error forecast correction. Energy Convers Manage 2016;119:215–26.
[116] 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.
[117] Wang Y, Wang HB, Srinivasan D, Hu QH. Robust functional regression for wind speed forecasting based on Sparse Bayesian learning. Renew Energy 2019;132:43–60.
[118] Hu Q, Zhang S, Xie Z, Mi J, Wan J. Noise model based ν-support vector regression with its application to short-term wind speed forecasting. Neural Networks 2014;57:1–11.
[119] Zhang S, Hu Q, Xie Z, Mi J. Kernel ridge regression for general noise model with its application. Neurocomputing 2015;149:836–46.
[120] Zuluaga CD, Alvarez MA, Giraldo E. Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison. Appl Energy 2015;156:321–30.
[121] Zheng L, Hu W, Min Y. Raw wind data preprocessing: a data-mining approach. Ieee T Sustain Energ 2015;6(1):11–9.
[122] Wang J, Xiong S. A hybrid forecasting model based on outlier detection and fuzzy time series–A case study on Hainan wind farm of China. Energy 2014;76:526–41.
[123] Wang J, Qin S, Zhou Q, Jiang H. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China. Renew Energy 2015;76:91–101.
[124] Gallego C, Pinson P, Madsen H, Costa A, Cuerva A. Influence of local wind speed and direction on wind power dynamics–Application to offshore very short-term forecasting. Appl Energy 2011;88(11):4087–96.
[125] Camelo HD, Lucio PS, Leal JBV, de Carvalho PCM. A hybrid model based on time series models and neural network for forecasting wind speed in the Brazilian northeast region. Sustain Energy Techn 2018;28:65–72.
[126] De Giorgi MG, Ficarella A, Tarantino M. Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy 2011;36(7):3968–78.
[127] Galicia A, Torres J, Martínez-Álvarez F, Troncoso A. A novel Spark-based multi-step forecasting algorithm for big data time series. Inf Sci 2018;467:800–18.
[128] Gensler A, Henze J, Sick B, Raabe N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. 2016 IEEE international conference on systems, man, and cybernetics (SMC). 2016. p. 002858–65.