IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2020, accepted March 25, 2020, date of publication March 30, 2020, date of current version April 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984020

Hybrid Prediction Method for Wind Speed Combining Ensemble Empirical Mode Decomposition and Bayesian Ridge Regression

YUJUN YANG[1,2,3] **AND YIMEI YANG**[1,2]

[1]School of Computer Science and Engineering, Huaihua University, Huaihua 418008, China
[2]Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, Huaihua 418000, China
[3]Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, Huaihua 418000, China

Corresponding author: Yimei Yang (yym1630@163.com)

This work was supported in part by the Scientific Research Fund of Hunan Provincial Education under Grant 19C1472 and Grant 17C1266, in part by the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province under Grant ZNKZ2018-5, in part by the Key Scientific Research Projects of Huaihua University under Grant HHUY2019-08, in part by the Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, and in part by the Constructing Program of the Key Discipline in Huaihua University.

ABSTRACT In recent years, with the rapid development of wind power generation, some problems are gradually highlighted. At present, one of the essential methods to solve these problems is to predict wind speed. In this paper, a hybrid BRR-EEMD method is proposed for short-term wind speed prediction based on the Bayesian ridge regression prediction method and ensemble empirical mode decomposition. We use ensemble empirical mode decomposition of the hybrid method to decompose complex time series of wind speed into several relatively milder, more regular, and stable subsequences. Then each subsequence is carried out by using the Bayesian ridge regression method. The value of each subsequence is predicted by it. Finally, the value of multiple subsequences is fused to form the prediction results of the original complex time series of wind speed. In order to verify the proposed method comprehensively, this paper selects two data to test. According to the results, predicted values have shown higher accuracy compared with the various prediction methods. Therefore, the hybrid BRR-EEMD method is accurate and effective in predicting wind speed, which has practical significance and potential value.

INDEX TERMS Ensemble empirical mode decomposition, short-term predicting, Bayesian ridge regression, wind speed, time series.

I. INTRODUCTION

With the growth of the global population and the rapid development of the global economy, economic development and the progress of science and technological progress are proliferating, and the demand for energy is multiplying. The increasing energy demand has led to the depletion of traditional fossil energy. Energy shortage, climate warming, and environmental pollution are becoming more and more dangerous. However, wind energy is free and clean. Wind energy has attracted more and more attention and has become one of the most popular renewable energy. In 2018, the new installed capacity of global wind power was 53.9 GW, and the cumulative installed capacity is more than 600 GW for the first time. In China, the new installed capacity is 21 GW, and the cumulative installed capacity is 221 GW [1]. According to data released by Bloomberg New Energy Finance, the global onshore wind power installed capacity in 2019 was 53.2GW, and the new installed capacity of offshore wind power reached a historical high of 7.5GW. China occupies half of the global market. In 2019, the new installed capacity was 28.9GW, including 26.2GW of onshore wind power and 2.7GW of offshore wind power, accounting for 48% of the global market. Accurate wind speed prediction plays an

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

essential role in the effective generation and distribution of wind power. So far, the researchers [2]–[14] proposed many wind speed prediction models [45]–[59].

Although the factors of changing wind speed are complex and random, this does not mean that wind speed is unpredictable. In recent years, many researchers and scholars have paid close attention to the analysis and prediction of wind speed. This kind of work can be summarized into two directions; one is deterministic method; the other is statistical method. The first method focuses on the physical theory of atmospheric and meteorological processes. It is a prediction method based on the physical model. This method requires the establishment of thermodynamic and dynamic equations to describe the evolution of the atmosphere. At the same time, we need to pay attention to much historical data. Therefore, this method usually uses specific mathematical equations to represent the diffusion model of wind speed. The trend of wind speed and wind energy can be predicted considering the boundary conditions and the actual terrain. The prediction model constructed by this method is very complex, the calculation is complex, the amount of calculation is tremendous, and the result is susceptible to the wrong initial information [15]–[17]. The prediction method based on statistical theory has been widely used recently. Common statistical models include grey model (GM) [18], [19], autoregressive integrated moving average (ARIMA) [20], [21], support vector regression (SVR) [22], [23], multiple linear regression (MLR) [24], long-term and short-term memory network (LSTM) [9], [25], [26], artificial neural network(ANN) [27], machine learning algorithm(ML) [22], deep learning algorithm(DL) [29] and other hybrid models [30]. More and more researchers use machine learning methods to extract internal patterns from data. These methods were published in nature in 2019 [31]. The physical model based on theory and the machine learning model based on data have different characteristics. Nevertheless, the two methods can complement each other. The former has strong extrapolation ability. The latter is more flexible and can find new rules. For instance, Alexiadis *et al.* [32] proposed a technique for predicting wind speed within several hours ahead, based on cross-correlation at neighboring sites. Meanwhile, the authors develop an artificial neural network that significantly improves predicting accuracy comparing to the persistence predicting methods. Barbounis *et al.* [33] proposed two novel and optimal on-line learning schemes for the update of the recurrent network's weights using the recursive prediction error algorithm. The proposed methods assure the continuous stability of the network during the learning phase and exhibit improved performance compared to the conventional dynamic backpropagation. Simulation results demonstrate that the recurrent models, trained by the suggested methods, outperform the static ones while they exhibit significant improvement over the persistent method. Azad *et al.* [14] proposed a method of long-term wind speed predicting and general pattern recognition using neural networks. It may be a solution regarding the long-term wind speed forecast in order to solve the two problems, which are inefficient and less reliable results on account of wind speed with unstable and intermittent characteristics. Zhang *et al.* [34] proposed a sophisticated deep-learning method for short-term and long-term wind speed forecasts. The sophisticated deep-learning method forecasts wind speed by analyzing the higher level features abstracted from lower level features of the wind speed data. These automatically learned features are very informative and appropriate for the wind speed forecast. The sophisticated deep-learning method is a deep stochastic model that can represent the wind speed very well. The evaluation of the sophisticated deep-learning method is depicted by both hour-ahead and day-ahead prediction experiments based on real wind speed data in the world. The prediction accuracy of the sophisticated deep-learning method outperforms existing methods. Ren *et al.* [35] proposed a novel wind speed predicting method by integrating empirical mode decomposition and support vector regression methods. The EMD-SVR wind speed predicting method is evaluated by a wind speed data set. The EMD-SVR wind speed predicting method outperforms several recently reported methods concerning the accuracy or computational complexity. Khodayar *et al.* [36] proposed a deep neural network architecture with stacked autoencoder and stacked denoising autoencoder for ultrashort-term and short-term wind speed predicting. In order to improve the accuracy of current methodologies, rough neural networks are incorporated in the deep neural network architecture with stacked autoencoder and stacked denoising autoencoder to develop novel rough extensions of SAE and SDAE which are robust for wind uncertainties. Experimental results show that the deep neural network architecture with stacked autoencoder and stacked denoising autoencoder outperform classic DNNs and previous methods in the view of lower RMSE and mean absolute error measurements. Zhang *et al.* [37] proposed a wind speed predicting using a two-stage predicting system with an error-correcting and nonlinear ensemble strategy. The two-stage predicting system is performed based on the data preprocessing approach, improved multi-objective optimization algorithm, error correction, and nonlinear ensemble strategy. The two-stage predicting system effectively overcomes the shortcomings of wind speed predicting capacity, so it is more conducive to enhancing predicting precision and stability than other involved methods. All the methods mentioned above focus on the prediction values of wind speed using a limited data set. On the contrary, how to effectively decompose the time series data of wind speed to make prediction becomes a challenge. Due to the uncertainty and diffusion of wind speed, it is not enough that some particular statistical methods tend to introduce the biases of wind speed prediction. However, hybrid methods can get better prediction results than some particular statistical methods. Besides, the ensemble empirical mode decomposition method can divide and conquer of complex problems. Therefore, based on the Bayesian ridge regression prediction method and ensemble empirical mode decomposition method, the

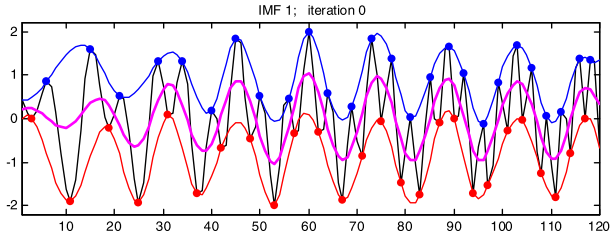

FIGURE 1. The decomposition process chart of sequences.

proposed hybrid BRR-EEMD method is used to predict the wind speed.

The rest of this paper is arranged as follows. Section 2 mainly introduces the related concepts and terms, for example, the EMD, the EEMD, and the BRR. The third part describes the comprehensive data collection, preprocessing, and modeling process. The experimental results of the proposed method are described in Section 4. Section 5 discusses the simulation experiment and empirical analyzes of the proposed hybrid method for prediction. Finally, the last section is the conclusion and some future works of this paper.

II. RELATED WORKS

The related concepts and terms, such as Empirical Mode Decomposition (EMD) method, Ensemble Empirical Mode Decomposition (EEMD) method, and Bayesian Ridge Regression (BRR) method related to this paper, will be introduced as follows.

A. EMPIRICAL MODE DECOMPOSITION

Huang *et al.* proposed the empirical mode decomposition (EMD) [38] in 1998, which is a self-adaptive time-frequency analysis method and a very effective time series decomposition method. Based on the local characteristics of time series data, the empirical mode decomposition can effectively extract the original time-series data from the time series with noise. It can also achieve better results in the decomposition of complex time series. Therefore, it has been successfully applied in many practical fields. The EMD has three assumptions which are shown as following [39]:

(1) The target signal must have at least two extrema, one maximum, and one minimum.

(2) The characteristic time scale is defined by the time lapse between the extrema.

(3) If there is no extreme value but only inflection point in the data, they can be distinguished one or more times to reveal the extrema.

The decomposition process of EMD is shown as follows

(1) For a given time series $x(t)$ is shown in the black line in Fig.1, all local extremum points, which are shown in the blue points and red points in Fig. 1, are obtained, and their upper and lower envelope lines, which are shown in the blue line and red line in Fig. 1, are obtained by using a cubic spline.

(2) The mean value $m(t)$ of the upper and lower envelopes, which is shown in the thick purple-red line in Fig.1, and then the difference between original signal $x(t)$ and mean value $m(t)$ defined $h(t)$ as Eq.1.

$$h(t) = x(t) - m(t) \tag{1}$$

(3) Determine whether $h(t)$ is an IMF. If $h(t)$ is an IMF, the result of Eq.1 is defined as the ith IMF $c\,(i)$. Otherwise, $h(t)$ is regarded as the original signal $x(t)$, and the above steps are repeated until $h(t)$ becomes a real IMF. After this, the ith IMF $c\,(i)$ is decomposed by

$$c\,(i) = h\,(i)\,, \quad i = 1, 2, \ldots, n-1 \tag{2}$$

(4) Separate the ith IMF $c\,(i)$ from $x(t)$ by

$$r\,(i) = x(t) - c\,(i) \tag{3}$$

where $r\,(i)$ is defined as the residue signal.

(5) Repeat the above steps K times until the stop condition takes place. Then, K IMFs will be obtained, and they satisfy

$$\begin{cases} r_1 - c_2 = r_2 \\ \quad \vdots \\ r_{K-1} - c_K = r_K \end{cases} \tag{4}$$

Finally, the original signal $x(t)$ can be decomposed into

$$x(t) = \sum_{k=1}^{K} c_k(t) + r_K(t) \tag{5}$$

B. ENSEMBLE EMPIRICAL MODE DECOMPOSITION

Traditional EMD has the problem of mode mixing when analyzing vibration signals. To solve this problem, many researchers put forward some solutions. In 2009, Wu and Huang [40] proposed an adaptive empirical mode decomposition method, which is a new, improved version of traditional EMD, named EEMD. The EEMD is a non-stationary signal analysis method, but it is different from the Fast Fourier Transform and Wavelet Transform. The EEMD is suitable for any data, and it can decompose data completely without the need for base function based on the data itself. The key to this method is empirical mode decomposition, which can decompose complex signals into finite intrinsic mode functions (IMF). Each IMF component decomposed contains local characteristic signals of different time scales of the original signal. The empirical mode decomposition method can make the non-stationary data to be stabilized, and then Hilbert transform can be used to obtain the time-frequency spectrum, which is of physical significance. Compared with Fast Fourier Transform and Wavelet Transform decomposition, this method is intuitive, direct, posterior, and adaptive, because the data itself decomposes the basic function. Because the decomposition is based on the local characteristics of the time scale of the signal sequence, it has the adaptive characteristics. Here is a brief introduction to the process of decomposing data with the EEMD method.

Suppose the EEMD method decomposes the sequences X into n subsequences. According to the decomposition regulation of the EEMD method, the last subsequence is called residual subsequence R. In fact, the EEMD method decomposes the sequences X into $n-1$ subsequence

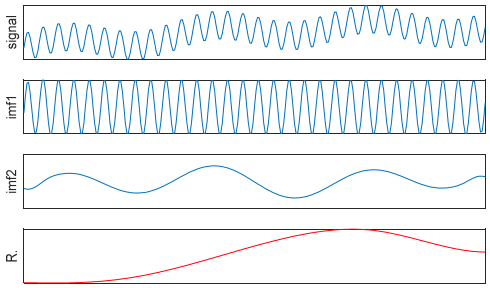

FIGURE 2. The EEMD decomposition process chart of Sin sequences.

component $C_i(i = 1, 2, \ldots, n-1)$, named $\mathrm{IMF_i}(i = 1, 2, \ldots, n-1)$, and a residual subsequence R. The following is a brief introduction to the steps of the EEMD method decomposition process.

(1) For a given sequence $x(t)$ is shown in the black line in Fig.1, the cubic spline interpolation method of local maximum and minimum values is used to create its upper and lower envelope lines, which are shown in the blue line and red line in Fig. 1.

(2) Calculate the mean value $m(t)$ of the upper and lower envelope lines, as shown in the thick purple-red line in Fig.1.

(3) By subtracting the mean value $m(t)$ of the upper and lower envelope from the original sequences $x(t)$, the first subsequence $\mathrm{IMF_1}$ component $h(t) = x(t) - m(t)$ is obtained.

(4) Taking the subsequence component $h(t)$ as a new sequence $x(t)$, and repeating steps (1) to (3) until the stop condition is met. The stop conditions are as follows:

(a) The mean value $m(t)$ of the upper and lower envelope lines is approximately equal to zero;

(b) The number of extreme points of component $h(t)$ is equal to or at most different from the number of zero-crossing points;

(c) The predefined maximum number of iterations has been reached.

(5) The subsequence component $h(t)$ is taken as a subsequence IMF component $\mathrm{C}_i(i = 1, 2, \ldots, n-1)$, and calculate the remaining subsequence R: $R(t) = x(t) - h(t)$.

(6) Use the remaining subsequence $R(t)$ as the new sequences $x(t)$ to calculate the next IMF, and repeat steps (1) to (5) until all the IMF is obtained or the maximum decomposition level is reached.

Finally, the EEMD method has decomposed completely sequences X into n subsequences, which are expressed by Eq.6.

$$x(t) = \sum_{i=1}^{n-1} (C_i) + R_n, \quad i = 1, 2, \ldots, n-1 \tag{6}$$

where the number n of subsequences depends on the complexity of the original sequences. Fig. 2 shows a decomposition process of the EEMD for the time series Eq.7.

$$x(t) = sin(2\pi * 10t) + 2 * sin(2\pi * 100t) + 3 * sin(2\pi * t) \tag{7}$$

where $t = 0, f, 2f, \ldots, 300f, f = 0.001$.

C. BAYESIAN RIDGE REGRESSION

Because Bayesian ridge regression is a particular case of Bayesian linear regression and belongs to ridge regression, it has all the characteristics of ridge regression and Bayesian linear regression. We first introduce the Bayesian estimation and the Bayesian linear regression [41].

Define a set of samples as D, the samples in the sample set D are all extracted independently from a fixed but unknown probability density function $p(x)$. It is required to estimate the probability distribution of x based on these samples, which is recorded as $p(x|D)$. Moreover make $p(x|D)$ as close as possible to $p(x)$. This is the core of Bayesian estimation. Although $p(x)$ is unknown, it is well known that a density distribution has two elements: form and parameter. We can assume that the form of $p(x)$ is known, but the value of parameter θ is unknown. Here the $p(x|\theta)$ is the first important element of Bayesian estimation, which also is a conditional probability density function. Because the form of $p(x|\theta)$ is known, and the value of parameter θ is unknown, where x can be regarded as a test sample. So the conditional density function $p(x|\theta)$, in essence, is the likelihood estimation of θ at point x. Since the value of parameter θ is unknown, θ can be regarded as a random variable. Then, a prior probability density function $p(\theta)$ can be used to represent the training samples before they are observed. By observing the training samples, we can transform the prior probability density into the posterior probability density function $p(\theta|D)$. According to the posterior probability density correlation, we hope that $p(\theta|D)$ has a very significant peak near the real value of θ. This is the posterior probability density, the second major element of Bayesian estimation. Now, the core problem of Bayesian estimation $p(x|D)$ is connected with two important elements of Bayesian estimation: $p(x|\theta)$ and $p(\theta|D)$.

$$p(x|D) = \int p(x, \theta|D)d\theta = \int p(x|\theta, D)p(\theta|D)d\theta \tag{8}$$

In the above formula, x is the test sample, D is the training set, and the selection of x and D is independent. Therefore, $p(x|\theta, D)$ can be written as $p(x|\theta)$. Therefore, the core problem of Bayesian estimation is the following formula:

$$p(x|D) = \int p(x|\theta)p(\theta|D)d\theta \tag{9}$$

Here $p(x|\theta)$ is the likelihood estimate of θ on the test sample x, while $p(\theta|D)$ is the posterior probability of θ on the existing sample set. The posterior probability $p(\theta|D)$ is shown in the following formula.

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \tag{10}$$

$$p(D|\theta) = \prod_{k=1}^{n} p(x_k|\theta) \tag{11}$$

In order to clearly express that there are n samples in the sample set D, the notation is used here:

$$D^n = \{x_1, x_2, \ldots, x_n\} \tag{12}$$

According to the former formula, there are the following formulas when $n > 1$:

$$p(D^n|\theta) = p(x_n|\theta)p(D^{n-1}|\theta) \tag{13}$$

According to the previous formula, it is easy to get the following formulas:

$$p\left(\theta \mid D^n\right) = \frac{p\left(x_n \mid \theta\right) p\left(D^{n-1} \mid \theta\right) p\left(\theta\right)}{\int p\left(x_n \mid \theta\right) p\left(D^{n-1} \mid \theta\right) p\left(\theta\right) d\theta} = \frac{p(x_n|\theta)p(\theta|D^{n-1})}{\int p(x_n|\theta)p(\theta|D^{n-1})d\theta} \tag{14}$$

When there is no observed sample, $p(\theta|D^0) = p(\theta)$ is defined as the initial estimate of parameter $\theta\theta$. Then let the sample set enter the above formula in turn, we can get a series of probability density functions: $p\left(\theta \mid D^0\right), p\left(\theta \mid D^1\right), p\left(\theta \mid D^2\right), \ldots, p\left(\theta \mid D^n\right)$. This process is called the Bayesian recursive method of parameter estimation, also called incremental learning of Bayesian estimation.

According to the previous discussion on maximum likelihood estimation(MLE), it is easy to know that if MLE is applied to the linear regression model, the complexity of the model will be controlled by two factors: the number of basis functions and the number of samples. Although adding a regular term, which is a prior distribution of parameters to the MLE estimation, can limit the complexity of the model to a certain extent and prevent overfitting, the selection of the basis function still plays a decisive role in the performance of the model. Because MLE always makes the model too complex to produce the phenomenon of overfitting, the simple application of MLE is not particularly effective.

Cross-validation is an effective way to limit the complexity of the model and prevent overfitting, but it needs to divide the data into a training set and a test set, which is also a severe waste of data samples. Based on the above discussion, we can introduce Bayesian linear regression. Bayesian linear regression can not only solve the problem of overfitting in MLE but also make full use of data samples. Only using training samples can effectively and accurately determine the complexity of the model. The model faced here is a linear regression model, which is a linear combination of the basic functions of a group of input variables x. Its mathematical form is as follows:

$$y(x, w) = w_0 + \sum_{j=1}^{M} \omega_j \phi_j(x) \tag{15}$$

Here $\phi_j(x)$ is the basis function mentioned earlier. The total number of basis functions is M. If $\phi_0(x) = 1$ is defined, the above formula can be simply expressed as:

$$y(x, w) = \sum_{j=0}^{M} \omega_j \phi_j(x) = w^T \phi(x) \tag{16}$$

where, $w = (w_0, w_1, w_2, \ldots, w_M)$, $\phi = (\phi_0, \phi_1, \ldots, \phi_M)$

Then the probability of the linear model is expressed as follows:

$$p(t|x, w, \beta) = N(t|y(x, w), \beta^{-1} I) \tag{17}$$

Assuming that the parameter w satisfies the Gaussian distribution, the above formula is a prior distribution as follows.

$$p(w) = N(w|0, \alpha^{-1} I) \tag{18}$$

In general, we call $p(w)$ a conjugate prior. Here t is the target output corresponding to x. β^{-1} and α^{-1} correspond to the variance of the Gaussian distribution of the sample set and w, respectively. Here w is a parameter. Then, the logarithmic posterior probability function of the linear model as follows.

$$lnp\left(\theta \mid D\right) = lnp\left(w \mid T\right) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{\alpha}{2} w^T w + const \tag{19}$$

where T is the target value vector of the data sample, $T = \{t_1, t_2, \ldots, t_n\}$, $const$ is a constant quantity independent of the parameter T.

According to the above incremental learning of Bayesian estimation, the following formula can be easily obtained.

Here is a brief introduction to the Bayesian learning process: on the posterior probability $p(\theta|D^{n-1})$ of the previous training set D^{n-1}, multiply it by the likelihood estimation of the new test sample point x_n to get the posterior probability $p(\theta|D^n)$ of the new set D^n. Thus, it is equivalent to that $p(\theta|D^{n-1})$ becomes the prior probability distribution of $(\theta|D^n)$, which is shown as follows.

$$p(\theta|D^n) \propto p(x_n|\theta)p(\theta|D^{n-1}) \tag{20}$$

If the prior distribution of Bayesian linear regression is the following formula (21), the final posterior distribution formula becomes the following formula (22).

$$p(w) = N(w|0, \alpha^{-1} I) \tag{21}$$

$$lnp\left(\theta \mid D\right) = lnp\left(w \mid T\right) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{\alpha}{2} w^T w + const \tag{22}$$

III. METHODOLOGY

This chapter mainly introduces the structure and flowchart of the proposed hybrid BRR-EEMD predict method for wind speed based on the Bayesian ridge regression prediction method and ensemble empirical mode decomposition method. The proposed hybrid BRR-EEMD predict method use ensemble empirical mode decomposition method to decompose complex time series sequence into several relatively simple subsequences firstly. Bayesian ridge regression is used to predict the value of each subsequence. Finally, the prediction results of several subsequences are fused to form the prediction results of the original complex time series sequence. There are many standard fusion methods, such as addition, multiplication, function fusion, and so on.

The structure of our proposed hybrid BRR-EEMD predict method is shown in Fig.3. The main structure of our proposed hybrid predict method includes EEMD decomposition, BRR prediction, and fusion method. The proposed hybrid BRR-EEMD predict method is separate into five steps: collect data and preprocess data, decompose data to obtain several subsequences by EEMD method, forecast each subsequence one by one using BRR method. Finally, form the prediction results of the original complex time series sequence

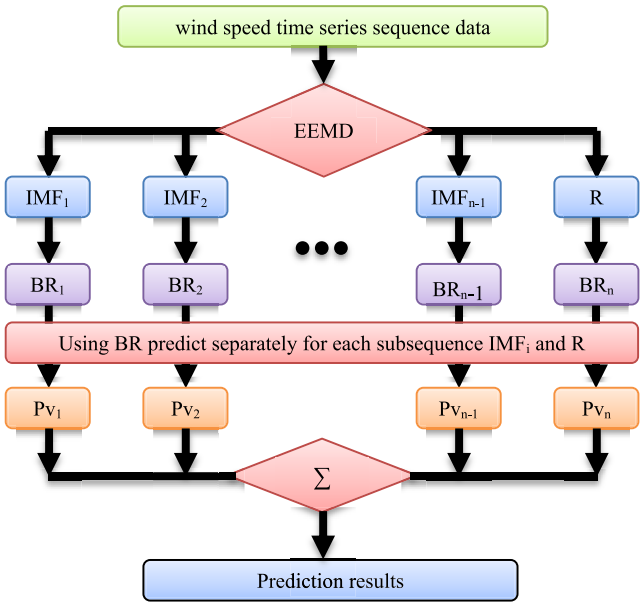

FIGURE 3. The structure of the proposed hybrid BRR-EEMD predict method.

fusing the prediction results of several subsequences. In this study, we choose addition as the fusion method. The following is a brief introduction to the structure of our proposed hybrid BRR-EEMD predict method.

(1) Generate simulation data and collect wind speed time series sequence data in the real world, preprocess the original wind speed data, make the data format of wind speed meet the format requirements of EEMD method decomposition, and form the input data X of hybrid BRR-EEMD predict method.

(2) EEMD method decomposes input data X into n subsequences. According to the regulation of the EEMD decomposition time series, the last subsequence is usually called residual subsequence R. Therefore, n subsequences include $n-1$ IMF subsequence and a residual subsequence R, which are respectively expressed as IMF_1, IMF_2, IMF_3, ..., IMF_{n-1}, R.

(3) Build one BRR model for each subsequence, and the processing of each subsequence is independent and does not affect each other, so n subsequences need to build n BRR models. After the n BRR models which are marked as $\mathrm{BRR_k}(\mathrm{k} = 1, 2, 3, \ldots, n-1, n)$ is built, predict of wind speed of n subsequences with n BRR models, n prediction results $\mathrm{Pv_k}(\mathrm{k} = 1, 2, 3, \ldots, n-1, n)$ are obtained accordingly.

(4) There are many fusion methods to fuse the prediction results of several subsequences to get the final prediction results of the original wind speed sequence. The addition is chosen as fusion method in this paper, which is to accumulate the prediction results of all subsequence to form the final prediction results.

(5) Finally, comparing the prediction results with the actual wind speed sequence, three evaluation criteria are used to calculate the prediction errors RMSE, MAE and R^2, and evaluate the advantages and disadvantages of hybrid BRR-EEMD predict method.

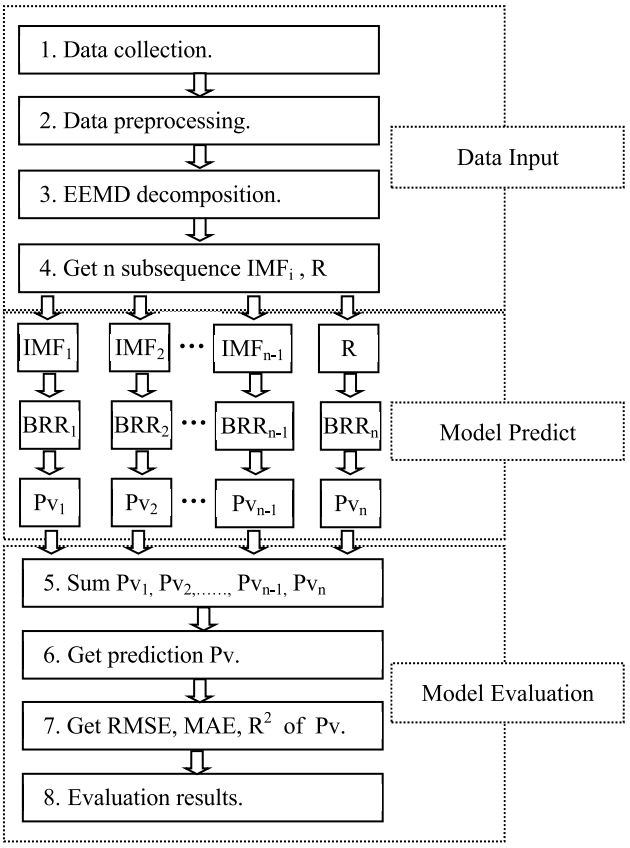

FIGURE 4. The flowchart of the proposed prediction method.

The flowchart of our proposed hybrid BRR-EEMD predict method is shown in Fig.4. The data flowchart of the proposed method is divided into three stages: data input, model prediction, and model evaluation. The data input stage includes data collection, data preprocessing, and data decomposition. In the prediction stage of the model, n subsequence data decomposed by EEMD are respectively predicted by the BRR model, and the prediction value of each subsequence data is obtained. The evaluation phase of the model includes the following steps. Firstly, the prediction value of n subsequence data is fused. In this paper, the fusion method of the sum is used, that is, sum the prediction value of n subsequence data, and then get the prediction value of the original sequence data. The next operation is to calculate the RMSE, MAE, and R^2 values of the prediction values in combination with the original sequence data. Finally, we use these values of RMSE, MAE, and R^2 to evaluate the model.

IV. EXPERIMENTS DATA

In this section, the experimental data are mainly introduced. In order to better test the prediction effect of our method, we use two kinds of experimental data. The first kind of experimental data is artificial simulation experimental data, which is automatically generated by computer according to the algorithm. This data is mainly used to verify the correctness and effectiveness of our proposed method. In many

literatures [42]–[44], artificial simulation data is also used to verify the effectiveness of the method. The second experimental data is the wind speed history data in the real world weather data. Only by using real data in the real world to test our proposed method is the most effective, so as to prove whether our method can be applied in the actual field.

A. ARTIFICIAL SIMULATION EXPERIMENTAL DATA

In the experiment, we use artificial simulation experiment data to test the validity and correctness of our method. In order to achieve the purpose of effectively testing our proposed method and obtaining sufficient and effective experimental results, the artificially generated simulation experimental data cannot be too small. Therefore, we choose 8000 artificial simulation data as experimental data. The artificial simulation data selected in the experiment is based on sin function composition, which is automatically generated by the computer program according to formula 23.

$$\begin{aligned} x(t) &= sin(2\pi * 100t) \\ y(t) &= sin(2\pi * 10t) \\ z(t) &= x(t) + y(t) \end{aligned} \tag{23}$$

where $t = 0, f, 2f, \ldots, 800f; f = 0.0001$ in Eq.23.

B. REAL SOCIAL EXPERIMENTAL DATA

In the experiment of this study, we use the actual data in the field of weather in real society to test the actual effect of our proposed method. The time series of wind speed in the meteorological field is used in this paper. The time series of wind speed in the field of Meteorology reflects the natural phenomena without human intervention. Moreover, the wind speed time-series fluctuates greatly, and the data are very representative, which can effectively test the effectiveness and practicability of our proposed method.

In the experiment, the wind speed time-series we use is the wind speed data of the air quality data set in the meteorological field. The wind speed time series data used in the experiment are weather and pollution conditions collected by the U.S. Embassy in Beijing in five years (January 1, 2010 to December 31, 2014). The data set is recorded in hours, including PM2.5 data, weather information, date and time, weather information, dew point, temperature, pressure, wind direction, wind speed, and accumulated snowfall hours. In the experiment, we only choose wind speed data for the experiment, because the change of wind speed time series is more intense than others. The data with more drastic changes can better test the validity and usability of our proposed method. The original data of the dataset used in the experiment can be downloaded from the UCI machine learning repository.

C. DATA PREPROCESSING

The data used in this experiment is time-series data, that is, continuous numerical data. In this paper, we use these data to establish a prediction problem. The prediction problem is to predict the value of the next time series according to the value of the 11 consecutive time series. That is to say, according to the 11 consecutive previous (hour or point) data of the above data, we predict the value of the 12^{th} (hour or point) data of the time series. The length of the experimental data is shown in Table 1. In Table 1, sin represents the artificially generated sin simulation time series, and wind speed represents the wind speed time series of the actual air quality data collected by the US embassy in Beijing.

TABLE 1. Length of data used in the experiment.

Category	Sin	Wind speed
Total length of selected original data	8000	8000
Total length of experimental data	7988	7988
Training data length	4792	4792
Test data length	3196	3196

In order to get better experimental results, we try to choose more data as experimental data and try to remove the invalid data or noise data. In practice, sensors are generally used to collect data, so there may be a fault in the sensor. In the case of sensor failure, the data collected is incorrect. Therefore, we need to remove invalid data or noise data before we start the experiment.

Because the prediction problem constructed in this study is to predict the value of the next series according to the value of the 11 consecutive previous sequences of the time series, it needs to preprocess the original wind speed time-series data. The original wind speed time series is stored in a one-dimensional array. The experiment needs to transform the original one-dimensional wind speed data into a two-dimensional array with 12 data as a group (row), which will result in the first part of the data not available. It causes the total length of the selected raw data (Row 1 in Table 1) to be 12 longer than the total length of the experimental data (Row 2 in Table 1). In order to ensure the comparability of the experimental results, we choose that the length of time series data of sin and wind speed is the same. In order to reduce the operation time, the first 8000 original data were selected. The length of the generated sin simulation time series is also 8000. Therefore, the ''Total length of selected original data'' of the two experimental data is displayed in the first row of Table 1, which is 8000. The ''Total length of experimental data'' of the two experimental data is shown in the second row of Table 1, which is 7988. In the experiment, the data need to be divided into test data and verification data. In the experiment, the ratio of 6:4 is used to segment the data. 60% of the total length of the experimental data is used as the test data, and the rest is used as the validation data. The specific length of test data and validation data of each time series is shown in the third and fourth rows of Table 1, respectively.

V. EXPERIMENTS

The experimental data used in the experiment come from two fields: one is the artificial simulation experimental data, the others are the real data in the real world. Besides, in order to compare with other methods conveniently, seven prediction methods are selected to test the same experimental data. Therefore, this section will explain and analyze the experimental results from three aspects.

A. EVALUATION CRITERION

There are many criteria for model evaluation. Most of the evaluation criteria are one-sided and targeted. For different problems, different evaluation standards must be used. Several typical problems are clustering, regression, classification, sorting, prediction, and recommendation. In order to evaluate our method correctly and effectively, we choose three evaluation criteria RMSE, R^2, and MAE, to calculate the prediction error and evaluate the proposed method.

The root mean square error (RMSE) is the mean square root of the sum of the squares of the distances that the data deviate from the real value, that is, the mean square root of the sum of the squares of the errors. The root mean square error is a widely used evaluation index and very sensitive to extremely small or large errors. The calculation formula is close to the standard deviation in form, which is shown in Eq.24.

$$RMSE\ (x, \widetilde{y}) = \sqrt{\frac{1}{m}\sum_{k=1}^{m}(x_k - \widetilde{y_k})^2} \tag{24}$$

where, x and $\widetilde{y}$ in Eq.24 represents the real value of time series and predicted value of time series respectively, x_k, and $\widetilde{y_k}$ represent the kth real value and the kth predicted value of time series respectively, and m represents the length of test data.

R-square (R^2) is the ratio of the sum of squares of the regression(SSR) to the sum of total deviation squares(STDS), which is called the coefficient of certainty. The SSR is the sum of the squares of the difference between the predicted data and the original data mean. The calculation formula of SSR is shown in formula 25. The STDS is the sum of the squares of the difference between the original data and the mean. The calculation formula of STDS is shown in formula 26. The calculation formula of R^2 is shown in formula 27.

The R^2 is to represent the fit of a model by the change of data. It can be seen from formula 27 that the normal value range of R^2 is [0.1]. The closer it is to 1, it indicates that the stronger the ability of the variables of the equation to interpret x is, the better the model fits the data.

$$SSR = \frac{1}{m}\sum_{k=1}^{m}(\widetilde{y_k} - \overline{\overline{y_k}})^2 \tag{25}$$

$$STDS = \frac{1}{m}\sum_{k=1}^{m}(x_k - \overline{\overline{y_k}})^2 \tag{26}$$

$$R^2\ (x, \widetilde{y}) = \frac{SSR}{STDS} = 1 - \frac{\sum_{k=1}^{m}(x_k - \widetilde{y_k})^2}{\sum_{k=1}^{m}(x_k - \overline{\overline{y_k}})^2} \tag{27}$$

In Eq.25, Eq.26, Eq.27, x, $\overline{\overline{y_k}}$, and $\widetilde{y}$ respectively represent the real value, the average value, and the predicted value of time series, m represents the length of time series, and the calculation formula of average value $\overline{\overline{y_k}}$ is shown in Eq.28.

$$\overline{\overline{y_k}} = \frac{1}{m}\sum_{k=1}^{m} y_k \tag{28}$$

Mean absolute error (MAE) is the average value of absolute error, which can better reflect the actual situation of predicted value error. The calculation formula of MAE is shown in Eq.29.

$$MAE\ (x, \widetilde{y}) = \frac{1}{m}\sum_{k=1}^{m}|x_k - \widetilde{y_k}| \tag{29}$$

where $x, \widetilde{y}, m$ in Eq.29 respectively represent the real value of time series, the predicted value of time series, and the length of time series.

B. ANALYSIS OF EXPERIMENTAL RESULTS OF OTHER METHODS

In order to compare with other prediction methods and find the effect and characteristics of the proposed method, six prediction methods are selected to predict the same experimental data. The experimental results are shown in Table 2. In order to let readers have a basic understanding of the six methods selected, the following is a brief introduction of these six methods.

1) NuSVR

NuSVR is short for Nu support vector regression. Nu support vector regression is similar to support vector regression. Nu support vector regression uses the parameter Nu to control the number of support vectors, which is implemented according to libsvm. In the experiment, the penalty parameter and nu parameter were 1 and 0.1, respectively, and no loss parameter was set.

2) GBR

GBR is short for Gradient boosting for regression. Gradient boosting for regression establishes an additive model in the forward stage, which allows any minor optimization of the loss function. In each stage, a regression tree is used to fit the negative gradient of a given loss function.

3) RFR

RFR is short for random forest regression. The random tree is a primary estimation method. The random tree method uses a large number of classification decision trees to fit the subsamples of data sets. Finally, the random tree method averages the fitting values of all subsamples to improve the prediction accuracy and fitting control.

4) KRR

KRR is short for kernel ridge regression. Kernel ridge regression combines the skill of ridge regression and kernel. At the same time, the kernel ridge regression is regularized by the

TABLE 2. Prediction results of three methods for three time-series.

METHOD		Sin	Wind speed
NuSVR	RMSE	0.033481	53.818728
	MAE	0.030063	48.212284
	R^2	0.998818	0.192938
GBR	RMSE	0.012665	19.688142
	MAE	0.009983	5.896139
	R^2	0.999831	0.840353
RFR	RMSE	0.005341	17.693007
	MAE	0.003571	6.105043
	R^2	0.999970	0.871070
KR	RMSE	0.838632	13.489963
	MAE	0.700675	5.650738
	R^2	0.258231	0.925050
KNR	RMSE	0.011126	19.522788
	MAE	0.005177	7.069213
	R^2	0.999869	0.843024
BRR	RMSE	**0.000035**	**13.382260**
	MAE	**0.000021**	**4.933050**
	R^2	**1.000000**	**0.926242**
Proposed BRR-EEMD	RMSE	**0.000033**	**11.928941**
	MAE	**0.000023**	**5.012087**
	R^2	**1.000000**	**0.941392**

linear least square L2 norm. In the data space of the online kernel, the kernel ridge uses the linear function in the original space. For the nonlinear kernel, kernel ridge regression uses a nonlinear function in the original space.

5) KNR

KNR is short for the K-nearest Neighbors Regression, which is a kind of regression method based on the k-nearest neighbor algorithm. The K-nearest neighbor regression method is easy to understand, and usually can get excellent performance without too much adjustment. The K-nearest neighbor regression method has a fast construction speed and is generally suitable for models with small training data sets. When the training data set of the K-nearest neighbor regression method is large; for example, the number of features and samples is large, the prediction speed may be slow. The K-nearest neighbor regression method has two important parameters: the number of neighbors and the distance between data points. In practice, the K-nearest neighbor regression method is best to set a small number of neighbors. Generally, when the parameter is set to 3-5, the K-nearest neighbor regression method often gets better results. However, in order to consider its performance comprehensively, it is necessary to adjust the parameter accordingly.

6) BRR

BRR is short for Bayesian ridge regression, which uses the regularized parameters λ (accuracy of weight) and α (accuracy of noise) to fit the model. Bayesian ridge regression is developed based on Bayesian linear regression. Bayesian ridge regression solves the problem that Bayesian linear regression is difficult to determine the model in Maximum likelihood estimation(MLE) by introducing the penalty parameters and the Bayesian method of ridge regression.

Table 2 shows the prediction results of seven methods of two time series: sin and wind speed. We use seven methods to predict two time-series data at the same time, respectively. The specific length of the used time-series data is shown in the second row of Table 1. In the experiment, the prediction problem is to predict the value of the next time series according to the value of 11 consecutive previous time series. The experimental results in Table 2 are calculated by RMSE, MAE and R^2 based on the predicted and real results. The smaller the results of RMSE, and MAE, the better the prediction effect. However, the closer the experimental result of R^2 is to 1, the more suitable the model is for the data, and the better the prediction effect of the prediction method is.

In order to compare the effect with other methods, we show the results of the BRR and our proposed BRR-EEMD method in bold, as shown in Table 2. It is found that the BRR and our proposed BRR-EEMD method have the best effect on the prediction of sin time series data. The KR method and the BRR method are better in predicting wind speed time-series data. Therefore, the BRR is the best method to predict the selected two time-series data.

By observing Table 2, it is easy to find that for the prediction of sin time series data, except for the KR method, the other six methods have a good prediction effect. By observing the prediction value in the first column of Table 2, it is found that the R^2 evaluation values of the other six methods are all greater than 0.99, which shows that the six methods have a better prediction effect on stable and regularly changing time series. The results show that the six methods are effective for the prediction of stable and regular time series. In order to show the prediction effect of these methods intuitively, we show them in the form of a graph. In order to reduce the number of pages, here we choose the prediction result graph of one of the six methods to display, as shown in Figure 5. In order to make the displayed result graph clear and distinguishable, only the last 300 predicted results are displayed in the result graph.

Observing the KR experimental results in Table 2, it is found that the prediction effect of this method for the time series data of wind speed is better than others. However, the prediction effect for the time series data of sin is very poor. Comparing the changes in wind speed and sin time series, we can find that the changes in sin time series are more regular and stable, while the changes in wind speed time series are more rapid than others. It shows that the KR method is more suitable for time series prediction with drastic changes. In Table 2, comparing the experimental results of the KR with that of the BRR method, it is found that the prediction effect of BRR method is better than others.

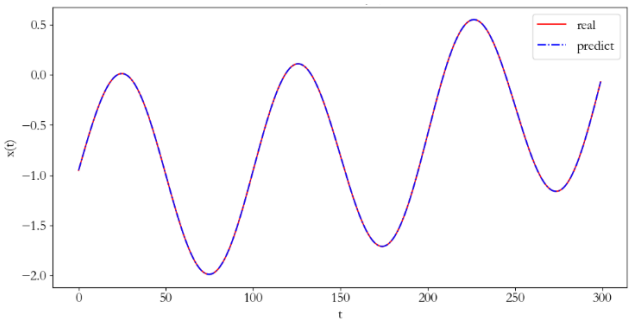

FIGURE 5. Prediction results of BRR method for sin time series.

By observing the prediction results of NuSVR in Table 2, it is easy to find that the prediction effect of this method on sin time series data is better than that on wind speed time-series data. NuSVR method is not effective in predicting the time series of wind speed with drastic changes. This shows that NuSVR is not suitable for time series prediction with irregular and unstable changes. According to table 2, only BRR and our proposed method BRR-EEMD can achieve excellent results for both wind speed time series and sin time series. In addition, other methods are not suitable for time series prediction with irregular and unstable.

C. ANALYSIS OF EXPERIMENTAL RESULTS BASED ON ARTIFICIAL SIMULATION DATA

In order to verify the correctness of our proposed method, we first use the artificial simulation experiment data of sin time series to test our method. In order to obtain sufficient and effective experimental results, the artificial simulation experimental data cannot be too small. However, in order to reduce the experimental time, the artificial simulation experimental data cannot be too much. Therefore, the length of sin simulation time series we selected is 8000.

Before discussing the experimental results of the actual wind speed data in the meteorological field, this paper first discusses the experimental results of forecasting the sin time series artificial simulation data by using our proposed BRR-EEMD method in this paper. The simulation experiment results verify the correctness and effectiveness of our proposed method.

Observing the sin data column of Table 2, the evaluation value of the BRR-EEMD method is equal to or better than that of the BRR method. RMSE and R^2 of the BRR-EEMD method are equal to or better than the BRR method. Especially on the RMSE index, the proposed method improves the experimental effect by 6.1% compared with the traditional method. Although the MAE evaluation value of the BRR-EEMD method is slightly lower than that of the BRR method, it can almost be ignored because the difference between them is tiny. In conclusion, our proposed BRR-EEMD prediction method has a better comprehensive effect than the BRR method, which shows that our proposed prediction method is correct and effective.

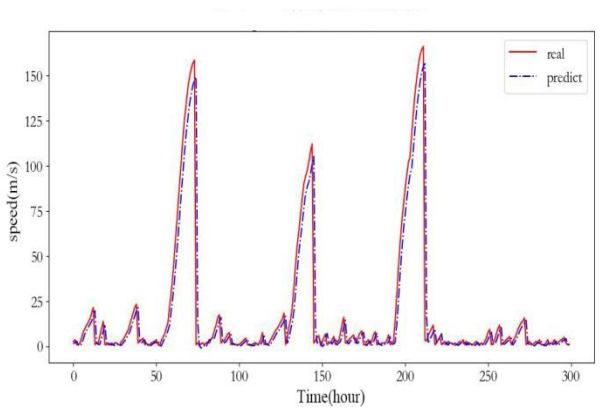

FIGURE 6. Prediction results of BRR-EEMD method for wind speed time series.

D. ANALYSIS OF EXPERIMENTAL RESULTS BASED ON REAL DATA

In order to verify the practicability of the proposed BRR-EEMD prediction method, we use the wind speed time series in the meteorological field in the experiment. The time series of wind speed is a natural phenomenon, which is not interfered with by human beings. However, the change of time series of wind speed is quite intense, and the sudden and sharp change occurs from time to time. The time series of wind speed has the characteristics of complexity, instability, and drastic change.

Observing the wind speed data column of Table 2, the evaluation value of the BRR-EEMD method is equal to or better than that of the BRR method. RMSE and R^2 of the BRR-EEMD method are equal to or better than the BRR method. On the RMSE and R^2 index, the proposed method improves the experimental effect by 12.18% and 1.6% compared with the traditional method, respectively. Although the MAE evaluation value of the BRR-EEMD method is slightly lower than that of the BRR method, it can almost be ignored because the difference between them is very small. In conclusion, the proposed BRR-EEMD prediction method has a better comprehensive effect than the BRR method, which shows that the proposed prediction method has good practical application value. Fig.6 shows the BRR-EEMD prediction results of wind speed time series.

By observing Table 2, it is easy to find that the prediction results of the BRR-EEMD proposed method in wind speed and sin time series data are better than that of the BRR method. There are two main reasons for such good experimental results. On the one hand, the proposed method decomposes the wind speed and sin time-series data through EEMD, which makes the complicated original wind speed and sin time series data decompose into more regular and stable subsequences. Moreover, the prediction results of subsequences are more accurate. On the other hand, although the changes in wind speed and sin time series are complex, they can be decomposed into many regular and stable time series using the proposed method. In order to clearly understand

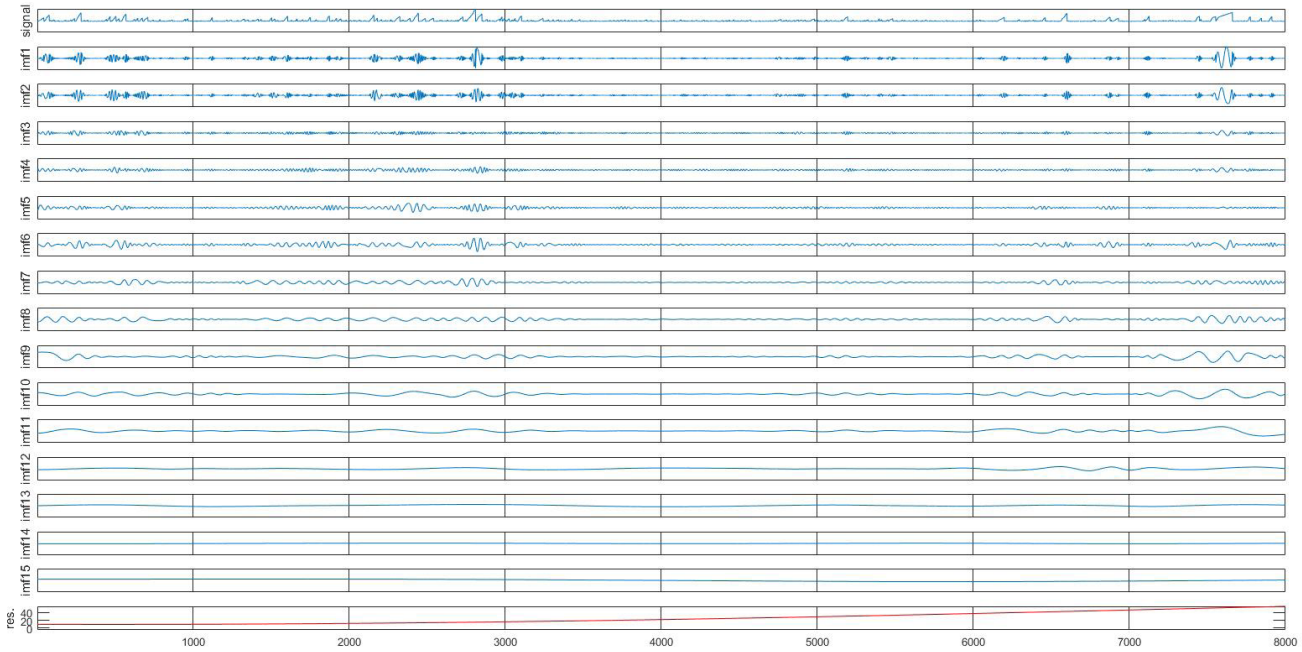

FIGURE 7. Wind speed time-series change and EEMD decomposition.

the change of wind speed time series and its EEMD decomposition results. Figure 7 shows the EEMD decomposition results of wind speed time-series data. The first subgraph at the top of Figure 7 is the fluctuation graph of the original wind speed time series, and the other subgraphs are EEMD decomposition subgraphs of wind speed time series. It is easy to see that the lower EEMD decomposition subgraph is simpler, more regular, and more stable. In theory, it is easier to predict for these simpler, more regular, and more stable subgraphs, and the prediction effect is better than that of the original.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a hybrid method for short-term time series forecasting based on EEMD and BRR. The method is based on the idea of divide and conquer of complex problems, combined with EEMD and Bayesian Ridge Regression. EEMD is used by the proposed BRR-EEMD prediction method to decompose complex time series into several relatively simpler, more regular, and more stable subsequences. Then the BRR method is used to predict the value of each subsequence. The prediction results of several subsequences are fused to form the prediction results of the original wind speed time series. In order to verify the effectiveness and practicability of the proposed method, two representative time series data are selected for testing in the experiment. Experimental results show that the proposed method has an excellent comprehensive effect.

The research of time series analysis and prediction method has developed rapidly. However, its effect cannot meet the high requirements of practical application in some fields in many aspects, and there are many problems to be solved. Based on the research work in this paper, there are still many contents to study. In the future, we intend to study many prediction methods and machine learning methods systematically. We hope to propose some new methods based on these methods and EEMD method and improve the prediction effect of time series, to apply them to the actual field correctly and effectively.

REFERENCES

[1] Y. Zhang, H. Sun, and Y. Guo, "Wind power prediction based on PSO-SVR and grey combination model," *IEEE Access*, vol. 7, pp. 136254–136267, 2019.

[2] M. R. Wani, M. A. Wani, and R. Riyaz, "Cluster based approach for mining patterns to predict wind speed," in *Proc. IEEE Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, Birmingham, U.K., Nov. 2016, pp. 1046–1050.

[3] Z. Yanning, K. Longyun, Z. Shiqiong, and C. Binggang, "Wind speed predicted by wavelet analysis in input prediction control of wind turbine," in *Proc. 7th World Congr. Intell. Control Autom.*, Jun. 2008, pp. 7732–7735.

[4] S. Yoshida, H. Suzuki, T. Kitajima, A. M. Kassim, and T. Yasuno, "Correction method of wind speed prediction system using predicted wind speed fluctuation," in *Proc. 55th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE)*, Tsukuba, Japan, Sep. 2016, pp. 1054–1059.

[5] W. Yao, P. Huang, and Z. Jia, "Multidimensional LSTM networks to predict wind speed," in *Proc. 37th Chin. Control Conf. (CCC)*, Wuhan, Jul. 2018, pp. 7493–7497.

[6] Y. Zhang, S. Gao, J. Han, and M. Ban, "Wind speed prediction research considering wind speed ramp and residual distribution," *IEEE Access*, vol. 7, pp. 131873–131887, 2019.

[7] Savita, M. A. Ansari, N. S. Pal, and H. Malik, "Wind speed and power prediction of prominent wind power potential states in India using GRNN," in *Proc. IEEE 1st Int. Conf. Power Electron., Intell. Control Energy Syst. (ICPEICES)*, Delhi, India, Jul. 2016, pp. 1–6.

[8] Q. Hu, P. Su, D. Yu, and J. Liu, "Pattern-based wind speed prediction based on generalized principal component analysis," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 866–874, Jul. 2014.

[9] Y.-X. Wu, Q.-B. Wu, and J.-Q. Zhu, "Data-driven wind speed forecasting using deep feature extraction and LSTM," *IET Renew. Power Gener.*, vol. 13, no. 12, pp. 2062–2069, Sep. 2019.

[10] S. J. Watson, L. Landberg, and J. A. Halliday, "Application of wind speed forecasting to the integration of wind energy into a large scale power system," *IEE Proc. Gener., Transmiss. Distrib.*, vol. 141, no. 4, pp. 357–362, Jul. 1994.

[11] M. Mao, J. Ling, L. Chang, N. D. Hatziargyriou, J. Zhang, and Y. Ding, "A novel short-term wind speed prediction based on MFEC," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 4, no. 4, pp. 1206–1216, Dec. 2016.

[12] T. H. M. El-Fouly, E. F. El-Saadany, and M. M. A. Salama, "One day ahead prediction of wind speed and direction," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 191–201, Mar. 2008.

[13] Q. Zhu, J. Chen, D. Shi, L. Zhu, X. Bai, X. Duan, and Y. Liu, "Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 509–523, Jan. 2020.

[14] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 546–553, Apr. 2014.

[15] J. Wang, J. Hu, K. Ma, and Y. Zhang, "A self-adaptive hybrid approach for wind speed forecasting," *Renew. Energy*, vol. 78, pp. 374–385, Jun. 2015.

[16] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renew. Sustain. Energy Rev.*, vol. 34, no. 6, pp. 243–254, Jun. 2014.

[17] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renew. Sustain. Energy Rev.*, vol. 31, no. 2, pp. 762–777, Mar. 2014.

[18] M. Guo, Z. Wei, H. Zang, G. Sun, H. Li, and K. W. Cheung, "Short term wind speed forecasting using wavelet transform and grey model improved by particle swarm optimization," in *Proc. 5th Int. Conf. Electric Utility Deregulation Restruct. Power Technol. (DRPT)*, Changsha, China, Nov. 2015, pp. 1879–1884.

[19] C.-Y. Huang, Y.-W. Liu, W.-C. Tzeng, and P.-Y. Wang, "Short term wind speed predictions by using the grey prediction model based forecast method," in *Proc. IEEE Green Technol. Conf. (IEEE-Green)*, Baton Rouge, LA, USA, Apr. 2011, pp. 1–5.

[20] V. Radziukynas and A. Klementavicius, "Short-term wind speed forecasting with ARIMA model," in *Proc. 55th Int. Sci. Conf. Power Electr. Eng. Riga Tech. Univ. (RTUCON)*, Riga, Latvia, Oct. 2014, pp. 145–149.

[21] E. Yatiyana, S. Rajakaruna, and A. Ghosh, "Wind speed and direction forecasting for wind power generation using ARIMA model," in *Proc. Australas. Universities Power Eng. Conf. (AUPEC)*, Melbourne, VIC, Australia, Nov. 2017, pp. 1–6.

[22] Q. Hu, S. Zhang, M. Yu, and Z. Xie, "Short-term wind speed or power forecasting with heteroscedastic support vector regression," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 241–249, Jan. 2016.

[23] N. Botha and C. M. van der Walt, "Forecasting wind speed using support vector regression and feature selection," in *Proc. Pattern Recognit. Assoc. South Afr. Robot. Mechatronics (PRASA-RobMech)*, Bloemfontein, South Africa, Nov. 2017, pp. 181–186.

[24] U. Sanusi and D. Corne, "Feature selection for accurate short-term forecasting of local wind-speed," in *Proc. IEEE 8th Int. Workshop Comput. Intell. Appl. (IWCIA)*, Hiroshima, Japan, Nov. 2015, pp. 121–126.

[25] P. P. P, V. Vanitha, and R. R, "Wind speed forecasting using long short term memory networks," in *Proc. 2nd Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT)*, Kannur, India, Jul. 2019, pp. 1310–1314.

[26] F. R. Ningsih, E. C. Djamal, and A. Najmurrakhman, "Wind speed forecasting using recurrent neural networks and long short term memory," in *Proc. 6th Int. Conf. Instrum., Control, Autom. (ICA)*, Bandung, Indonesia, Jul. 2019, pp. 137–141.

[27] D. Yuan, Z. Qian, and Y. Pei, "State transition ANNs for hourly wind speed forecasting," in *Proc. Chin. Autom. Congr. (CAC)*, Jinan, China, Oct. 2017, pp. 6934–6938.

[28] Q. Hu, S. Zhang, M. Yu, and Z. Xie, "Short-term wind speed or power forecasting with heteroscedastic support vector regression," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 241–249, Jan. 2016.

[29] M. Afrasiabi, M. Mohammadi, M. Rastegar, and A. Kargarian, "Probabilistic deep neural network price forecasting based on residential load and wind speed predictions," *IET Renew. Power Gener.*, vol. 13, no. 11, pp. 1840–1848, Aug. 2019.

[30] N. Tomin, D. Sidorov, V. Kurbatsky, V. Spiryaev, A. Zhukov, and P. Leahy, "A hybrid wind speed forecasting strategy based on Hilbert-huang transform and machine learning algorithms," in *Proc. Int. Conf. Power Syst. Technol.*, Chengdu, China, Oct. 2014, pp. 2980–2986.

[31] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019.

[32] M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *IEEE Trans. Energy Convers.*, vol. 14, no. 3, pp. 836–842, Sep. 1999.

[33] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp. 273–284, Mar. 2006.

[34] C.-Y. Zhang, C. L. P. Chen, M. Gan, and L. Chen, "Predictive deep Boltzmann machine for multiperiod wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1416–1425, Oct. 2015.

[35] Y. Ren, P. N. Suganthan, and N. Srikanth, "A novel empirical mode decomposition with support vector regression for wind speed forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1793–1798, Aug. 2016.

[36] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Dec. 2017.

[37] L. Zhang, Y. Dong, and J. Wang, "Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy," *IEEE Access*, vol. 7, pp. 176000–176023, 2019.

[38] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "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.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[39] F. Jiang, Z. Zhu, and W. Li, "An improved VMD with empirical mode decomposition and its application in incipient fault detection of rolling bearing," *IEEE Access*, vol. 6, pp. 44483–44493, 2018.

[40] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[41] M. Byrtek, F. O'Sullivan, M. Muzi, and A. M. Spence, "An adaptation of ridge regression for improved estimation of kinetic model parameters from PET studies," *IEEE Trans. Nucl. Sci.*, vol. 52, no. 1, pp. 63–68, Feb. 2005.

[42] Y. Yang, J. Li, and Y. Yang, "The cross-correlation analysis of multi property of stock markets based on MM-DFA," *Phys. A, Stat. Mech. Appl.*, vol. 481, pp. 23–33, Sep. 2017.

[43] X. J. Zhao, P. J. Shang, and Y. L. Pang, "Power law and stretched exponential effects of extreme events in Chinese stock markets," *Fluct. Noise Lett.*, vol. 9, no. 2, pp. 203–217, 2010.

[44] Y. Yujun, L. Jianping, and Y. Yimei, "Multiscale multifractal multiproperty analysis of financial time series based on Rényi entropy," *Int. J. Modern Phys. C*, vol. 28, no. 2, Feb. 2017, Art. no. 1750028.

[45] Y. Ren, P. N. Suganthan, and N. Srikanth, "Ensemble methods for wind and solar power forecasting—A state-of-the-art review," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 82–91, Oct. 2015.

[46] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, "A review on hybrid empirical mode decomposition models for wind speed and wind power prediction," *Energies*, vol. 12, no. 2, p. 254, 2019.

[47] Y. Huang, L. Yang, S. Liu, and G. Wang, "Multi-step wind speed forecasting based on ensemble empirical mode decomposition, long short term memory network and error correction strategy," *Energies*, vol. 12, no. 10, p. 1822, 2019.

[48] M. Santhosh, C. Venkaiah, and D. M. V. Kumar, "Short-term wind speed forecasting approach using ensemble empirical mode decomposition and deep Boltzmann machine," *Sustain. Energy, Grids Netw.*, vol. 19, Sep. 2019, Art. no. 100242.

[49] N. Bokde, A. Feijóo, and K. Kulat, "Analysis of differencing and decomposition preprocessing methods for wind speed prediction," *Appl. Soft Comput.*, vol. 71, pp. 926–938, Oct. 2018.

[50] Z. Tian, "Short-term wind speed prediction based on LMD and improved FA optimized combined kernel function LSSVM," *Eng. Appl. Artif. Intell.*, vol. 91, May 2020, Art. no. 103573, doi: 10.1016/j.engappai.2020.103573.
[51] Z. Tian, S. Li, and Y. Wang, "A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed," *Wind Energy*, vol. 23, no. 2, pp. 177–206, 2020.
[52] T. Zhongda, "Chaotic characteristic analysis of short-term wind speed time series with different time scales," *Energy Sources, A, Recovery, Utilization, Environ. Effects*, vol. 2019, no. 5, pp. 1–15, May 2019.
[53] Z. Tian, G. Wang, and Y. Ren, "Short-term wind speed forecasting based on autoregressivemoving average with echo state network compensation," *Wind Eng.*, vol. 44, no. 2, pp. 152–167, May 2019.
[54] Z. Tian, G. Wang, S. Li, Y. Wang, and X. Wang, "Artificial bee colony algorithm–optimized error minimized extreme learning machine and its application in short-term wind speed prediction," *Wind Eng.*, vol. 43, no. 3, pp. 263–276, Jun. 2019.
[55] Z. Tian, Y. Ren, and G. Wang, "Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM," *Energy Sources, A, Recovery, Utilization, Environ. Effects*, vol. 41, no. 1, pp. 26–46, 2019.
[56] Z. Tian, Y. Ren, and G. Wang, "An application of backtracking search optimization-based least squares support vector machine for prediction of short-term wind speed," *Wind Eng.*, vol. 44, no. 2, pp. 152–167, May 2019.
[57] T. Zhongda, R. Yi, and W. Gang, "Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine," *J. Elect. Eng. Technol.*, vol. 13, no. 5, pp. 1841–1851, 2018.
[58] Z. Tian, G. Wang, Y. Ren, S. Li, and Y. Wang, "An adaptive online sequential extreme learning machine for short-term wind speed prediction based on improved artificial bee colony algorithm," *Neural Netw. World*, vol. 28, no. 3, pp. 191–212, 2018.
[59] T. Zhongda, L. Shujiang, W. Yanhong, and S. Yi, "A prediction method based on wavelet transform and multiple models fusion for chaotic time series," *Chaos, Solitons Fractals*, vol. 98, pp. 158–172, May 2017.

YUJUN YANG received the B.S. and M.S. degrees in computer science and technology from the Central South University of Forestry Science and Technology, Changsha, China, in 2002 and 2007, respectively, and the Ph.D. degree in computer science and technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2018.

He is currently a Senior Research Professional with the Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, a Research Assistant with the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, and an Associate Professor with the School of Computer Science and Engineering, Huaihua University. His research interests include machine learning, time series analyzing, financial data mining, prediction and recommendation, and artificial intelligence techniques.

YIMEI YANG received the M.S. degree in computer science and technology from the Central South University of Forestry Science and Technology, Changsha, China, in 2007.

She is currently a Senior Research Professional with the Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, a Research Assistant with the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, and an Associate Professor with the School of Computer Science and Engineering, Huaihua University. Her research interests include machine learning, time series analyzing, financial data mining, prediction and recommendation, and artificial intelligence techniques.