*Research Article*

# Short-Term Wind Speed Forecasting Based on Ensemble Online Sequential Extreme Learning Machine and Bayesian Optimization

**Jicheng Quan** ⓘ **and Li Shang**

*Department of Computer Science and Technology, Tongji University, No. 4800 Caoan Rd, Jiading District, Shanghai 201804, China*

Correspondence should be addressed to Jicheng Quan; 1510485@tongji.edu.cn

Short-term wind speed forecasting is crucial to the utilization of wind energy, and it has been employed widely in turbine regulation, electricity market clearing, and preload sharing. However, the wind speed has inherent fluctuation, and accurate wind speed prediction is challenging. This paper aims to propose a hybrid forecasting approach of short-term wind speed based on a novel signal processing algorithm, a wrapper-based feature selection method, the state-of-art optimization algorithm, ensemble learning, and an efficient artificial neural network. Variational mode decomposition (VMD) is employed to decompose the original wind time-series into sublayer modes. The binary bat algorithm (BBA) is used to complete the feature selection. Bayesian optimization (BO) fine-tuned online sequential extreme learning machine (OSELM) is proposed to forecast the low-frequency sublayers of VMD. Bagging-based ensemble OSELM is proposed to forecast high-frequency sublayers of VMD. Two experiments were conducted on 10 min datasets from the National Renewable Energy Laboratory (NREL). The performances of the proposed model were compared with various representative models. The experimental results indicate that the proposed model has better accuracy than the comparison models. Among the thirteen models, the proposed VMD-BBA-EnsOSELM model can obtain the best prediction accuracy, and the mean absolute percent error (MAPE) is always less than 0.09.

## 1. Introduction

Wind energy has grown substantially for two decades [1]. It has become one of the primary renewable energy sources. However, wind energy is highly variable, which affects the stable operation of the grid. Wind speed prediction can enhance wind farm operations and reduce the influence of wind energy on the grid. As the installed capacity of wind energy increases year by year [2], the industry needs more accurate wind speed prediction, making this subject an essential topic in energy research. Over the past decade, scholars have proposed many wind speed prediction methods. These methods are divided into four categories, i.e., (1) physical methods, (2) statistical methods, (3) artificial intelligence methods, and (4) hybrid methods.

The physical methods are based on fluid dynamics principles to establish numeric weather prediction (NWP) models. These methods need vast calculations and are not suitable for short-term wind speed prediction [3]. Statistical methods can analyze the patterns in historical data and establish linear prediction models. Representative methods include autoregressive (AR) [4], autoregressive moving average (ARMA) [5], autoregressive integrated moving average (ARIMA) [6], and pattern sequence similarity (PSF) [7]. These methods are not capable of characterizing nonlinear relationships in the wind data to produce high-precision prediction results.

Artificial intelligence methods are good at modeling nonlinear relationships. Among the AI models, the most widely used ones are artificial neural networks (ANNs) [8] and support vector machines (SVMs) [9]. However, the ANNs have multilayer structures that contain many parameters to adjust. The SVM is sensitive to parameters and needs massive calculation on large data sets. Extreme

learning machine (ELM) is a simple neural network [10]. Compared to the ANNs, ELM has a single hidden layer and therefore has fewer network parameters. Compared to SVM, ELM is more efficient. Consequently, ELM is an excellent predictor [11]. For instance, Liu et al. [12] used the ELM to complete the forecasting for the high-frequency sublayers obtained by the VMD-SSA. Fu et al. [13] proposed a hybrid approach based on dominant ingredient chaotic analysis and the ELM. However, in these papers, ELM is in the offline mechanism, and they cannot support real-time learning. To address this issue, an online sequential extreme learning machine (OSELM) was introduced. Zhang et al. [14] proposed an online sequential outlier robust extreme learning machine (OSORELM) for short-term wind speed prediction. Tian et al. [15] proposed an adaptive OSELM to improve ELM's prediction ability further.

Ensemble learning, such as Bagging [16] and Boosting [17], can combine multiple weak predictors to complete the forecasting. Bagging can reduce the prediction variance and improve the stability of the fundamental predictors. Zontul et al. [18] proposed a Bagging-based decision tree algorithm for wind speed prediction. Emeksiz and Demir [19] used the Bagging algorithm to estimate wind speed. Boosting can effectively enhance the performance of a weak predictor. Peng et al. [17] used the AdaBoost neural network to solve the lower accuracy defect. Liu et al. [20] proposed an AdaBoost algorithm and the multilayer perceptron (MLP) neural networks.

Besides ensemble learning, hybrid methods can improve the prediction robustness and accuracy of a single model. In a hybrid model, signal decomposition algorithms are employed to reduce the prediction complexity. The representative algorithms are wavelet decomposition (WD), wavelet packet decomposition (WPD), empirical mode decomposition (EMD), and ensemble empirical mode decomposition (EEMD). For instance, Fei and He [21] proposed a hybrid prediction method that combined WD and relevance vector machine. Liu et al. [22] presented a novel approach based on WPD and convolutional long short-term memory (ConvLSTM) networks. Zhang et al. [23] developed a model combining EMD, ANN, and SVM. Tian et al. [24] proposed a prediction approach using EEMD and extreme learning machine (ELM). However, the above decomposition methods have shortcomings. For instance, the wavelet-based approaches do not support adaptive processing; EMD cannot avoid mode mixing, and EEMD can add extra white noise into the wind data. A novel signal processing method, variational mode decomposition (VMD), was proposed to overcome the above obstacles. It can break down the original wind speed time-series into a set of band-limited sublayer modes named intrinsic mode functions (IMFs). These IMFs are stationary to predict. For instance, Zhang et al. [25] presented a hybrid model of the ANN, VMD, and Lorenz disturbance. The results proved the stable prediction performance of the proposed model. Gendeel et al. [26] presented an ANN prediction model with VMD. The comparison results indicated that the proposed model obtained significant improvements in forecasting accuracy.

Feature selection methods can improve the computational efficiency of the hybrid models. The typical filter-based approaches are partial autocorrelation function (PACF) and information theory methods. Sun et al. [27] applied PACF to identify the correlation between the decomposed components of EEMD. Memarzadeh and Keynia [28] used mutual information (MI) for feature selection. Huang et al. [29] used conditional mutual information (CMI) to analyze the correlation between the input features. Compared with the filter methods, the metaheuristic optimization-based wrapper approach can produce better accuracy. Sun et al. [27] used the binary-value gravitation search algorithm (BGSA) to improve the regression performance. Liu et al. [30] used the binary-coded genetic algorithm (BGA) for feature selection. Recently, the binary bat algorithm (BBA) has been proposed for feature selection. Compared with other metaheuristic algorithms, BBA has fewer parameters to adjust and can obtain better accuracy. Naik et al. [31] used the BBA to identify the relevant subset of features for the machine-learning tasks. Xie et al. [32] applied BBA to realize test-cost-sensitive attribute reductions. Liu et al. [33] used BBA to remove redundant features for image steganalysis effectively. Since BBA is superior to PACF, it is employed for feature selection in this paper.

Besides the feature selection, the metaheuristic optimization algorithms can be used to seek the optimal parameters of the prediction models to promote the predictors' performance on the datasets [34]. Among the metaheuristic algorithms, the genetic algorithm (GA) [35] and particle swarm optimization (PSO) [36] have been widely used in wind speed prediction. Although they are suitable for optimizing the model parameters, they need massive calculations and are vulnerable to improper parameter initialization. In the past few years, Bayesian optimization (BO) has emerged as a powerful tool for fine-tuning hyperparameters. Specifically, BO is capable of optimizing expensive black-box objective functions. Compared with the evolutionary computation methods, BO can achieve desirable results with fewer iterations. For instance, Cho et al. [37] used BO to fine-tune deep neural networks. The experimental results indicated that BO is a robust solution compared to the existing solutions. Muhuri and Biswas [38] used BO to optimize task scheduling. Their approach obtained optimal schedules without violation of the constraints. The experimental results indicated that BO is sample-efficient and can significantly outperform existing optimizers.

This paper proposes a novel approach for short-term wind speed prediction based on the above issues. The proposed model combines VMD, BBA, OSELM, BO, and Bagging. The contributions of the paper are as follows:

(1) VMD is utilized to preprocess the original wind time-series into more stationary sublayers for prediction. Compared to EMD and its variants, the proposed approach is more robust to data noise

(2) BBA is employed to complete the feature selection. Compared to PACF, BBA can achieve better prediction accuracy

(3) BO-optimized OSELM, referred to as BO-OSELM, is used to forecast the low-frequency sublayers of VMD. Compared to ELM, OSELM can provide the capability of online learning. Besides, BO is used to optimize the structure of OSLEM

(4) Bagging-based ensemble OSELM, referred to as Bagging-OSELM, is employed to forecast high-frequency sublayers of VMD. The Bagging-OSELM reveals better stability and accuracy than OSELM and AdaBoost-OSELM

The remaining part of the paper proceeds as follows: Section 2 introduces the proposed hybrid model, Section 3 presents the experimental results and discussion, and Section 4 draws the conclusions.

## 2. The Proposed Hybrid Model

In this section, the proposed hybrid model, referred to as VMD-BBA-EnsOSELM, is presented. This approach combines VMD, BBA, BO, Bagging, and OSELM. The architecture of the proposed model is demonstrated in Figure 1. The process of the proposed method is introduced as follows:

(1) VMD is utilized to decompose the denoised original data set into stationary sublayers.

(2) The feature selection method of BBA is applied to reserve critical features from the sublayers produced by VMD. The past twenty data points of the wind speed are chosen as the candidate feature sets. BBA determines the most relevant features of the candidate features.

(3) OSELM is adopted to complete the forecasting for the low-frequency sublayers obtained by VMD. BO optimizes the parameters of OSELM.

(4) Bagging-OSELM is adopted to complete the forecasting for the high-frequency sublayers obtained by VMD.

(5) All the forecasting results of the BO-OSELM and Bagging-OSELM are aggregated to produce the final prediction results.

(6) The proposed model is evaluated and compared with twelve comparison models, including the GPR model, the LSSVR model, the LSTM model, the OSELM model, the AdaBoost-OSELM model, the Bagging-OSELM model, the BBA-OSELM model, the BO-OSELM model, the PSO-OSELM model, the EMD-BBA-OSELM model, the EEMD-BBA-OSELM model, and the VMD-BBA-OSELM model.

### 2.1. Variational Mode Decomposition. 

The VMD algorithm is developed to overcome the limitations of EMD [39]. It can decompose an original signal $x(t)$ into IMFs. In the literature, it presented significant advantages in time-series forecasting [40] and fault diagnosis [41]. The core principle of VMD is to realize the IMFs by resolving the constrained optimization problem as follows:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad (1)$$

subject to

$$\sum_{k=1}^{K} u_k(t) = x(t), \quad (2)$$

where $\{u_k\} = \{u_1, u_2, \ldots, u_K\}$ denotes the IMFs; $\{\omega_k\} = \{\omega_1, \omega_2, \ldots, \omega_K\}$ is a central frequency of each IMF in the Fourier frequency domain; and $\delta(t)$ represents a Dirac function. The constraint conditions are (1) the original signal equals the sum of all the IMFs; and (2) the sum of the modal bandwidths is the least. Moreover, a Lagrange multiplier is introduced as

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2$$
$$+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2$$
$$+ \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle, \quad (3)$$

where $\alpha$ denotes a penalty factor, guaranteeing the decomposition precision, and $\lambda$ is a Lagrangian multiplier to assure the rigidity of the constraint conditions. The optimal solution to the above optimization problem is achieved as follows:

$$\widehat{u}_k^{n+1}(\omega) = \frac{\widehat{f}(\omega) \sum_{i=1}^{k-1} \widehat{u}_i^{n+1}(\omega) - \sum_{i=k+1}^{K} \widehat{u}_i^n(\omega) + \left( \widehat{\lambda}^n(\omega)/2 \right)}{1 + 2\alpha \left( \omega - \omega_k^n \right)^2},$$
$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \widehat{u}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \widehat{u}_k(\omega) \right|^2 d\omega}, \quad (4)$$

where $u_k(t)$ is an IMF, $\widehat{u}_k^{n+1}(\omega)$ is the Fourier transform of $u_k(t)$, and $n$ denotes the number of iterations to resolve the problem.

### 2.2. Binary Bat Algorithm. 

Inspired by bats, a novel metaheuristic algorithm, named the bat algorithm [42], is developed. In this algorithm, each bat can use echolocation to detect prey. In each iteration, a bat $b_i$ actively adjusts the loudness $A_i$ and the rate of pulse emission $r_i$ according to the prey's distance. Firstly, each bat $b_i$ is initialized with the position $x_i$, the velocity $v_i$, and the frequency $f_i$. Then, for each iteration $t$, the bat $b_i$ can be updated according to the following equations:

$$f_i = f_{\min} + (f_{\max} - f_{\min})\beta, \quad (5)$$

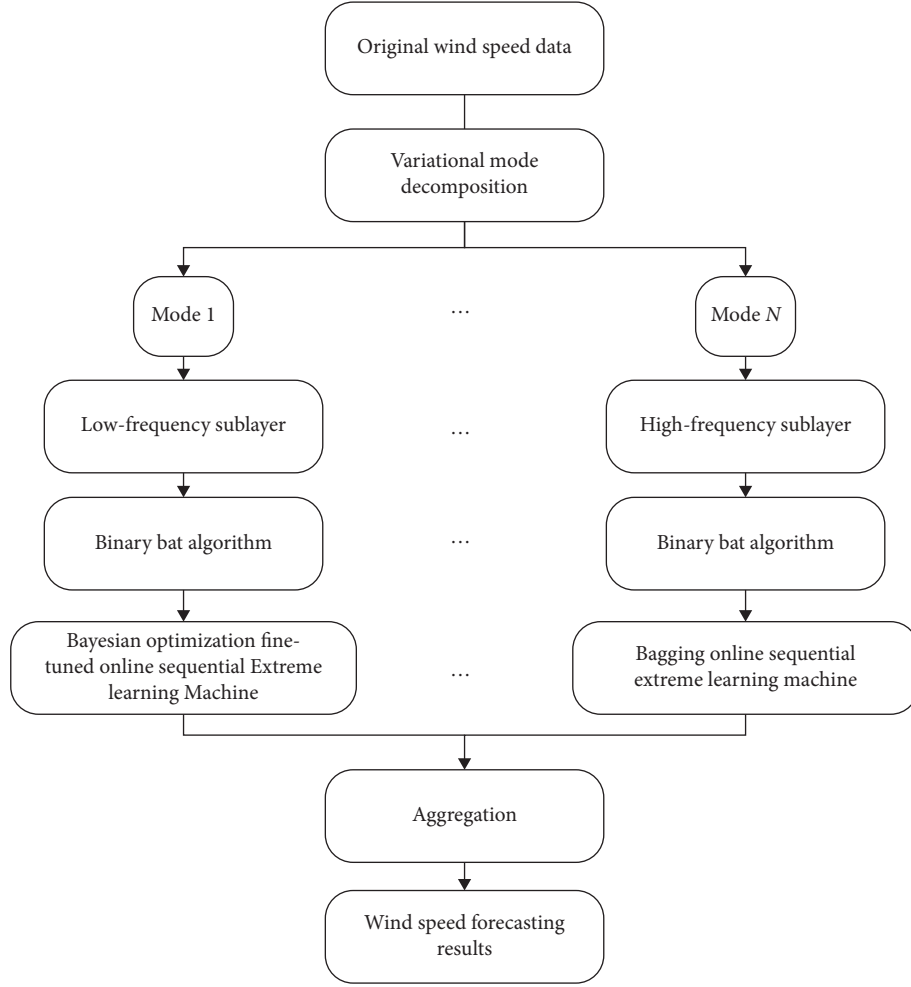$$v_i^j(t) = v_i^j(t-1) + \left[ \hat{x}^j - x_i^j(t-1) \right] f_i, \quad (6)$$

FIGURE 1: The process of the VMD-BBA-EnsOSELM model.

$$x_i^j(t) = x_i^j(t-1) + v_i^j(t), \tag{7}$$

$$A_i(t+1) = \alpha A_i(t), \tag{8}$$

$$r_i(t+1) = r_i(0) + \left[1 - e^{-\gamma t}\right], \tag{9}$$

where $\beta$ denotes a randomly generated number; $x_i^j(t)$ denotes the value of decision variable $j$ for bat $i$ at time step $t$; $\grave{x}^j$ represents the current global best solution for decision variable $j$; and $\alpha$ and $\gamma$ are user-specified constants (Algorithm 1).

In case of feature selection, a binary version of the bat algorithm is proposed restricting the new bat's position $x_i^j$ can be calculated as follows:

$$S\left(v_i^j\right) = \frac{1}{\left(1 + e^{-\left(v_i^j\right)}\right)},$$

$$\delta \sim U(0,1), \tag{10}$$

$$x_i^j = \begin{cases} -1, & \text{if } S\left(v_i^j\right) > \delta, \\ -0, & \text{otherwise.} \end{cases}$$

### 2.3. Online Sequential Extreme Learning Machine.
ELM is a novel feedforward network with a single hidden layer. The mathematical expression of ELM is illustrated as follows:

$$f(x) = G(x)\beta, \tag{11}$$

where $\beta$ is the output weight vector between the single hidden layer and the output layer, and $G(x)$ is the hidden layer output matrix. The optimal solution of $\beta$ can be obtained by

$$\beta^* = G^\dagger T, \tag{12}$$

where $G^\dagger$ is the Moore–Penrose inverse of $G$, and $\mathbf{T}$ is the training-target matrix.

OSELM is a novel online learning algorithm [43]. The algorithm can be divided into two phases: the initialization phase and the online learning phase. In the initialization phase, given a training dataset $T$, the hidden layer output matrix $G_0$ and the output weight vector $\beta_0^*$ can be calculated as follows:

$$\begin{aligned} N_0 &= \left(G_0^T G_0\right)^{-1}, \\ \beta_0^* &= N_0 G_0^T T. \end{aligned} \tag{13}$$

Then, the online learning process starts, and the algorithm learns the data block by block. In the $k$th iteration, a batch of new observed training-target matrix $T_0$ was given. The output weight vector $\beta_k^*$ can be calculated as follows:

$$N_k = N_{k-1} - \frac{N_{k-1} G_k G_k^T N_{k-1}}{1 + G_k^T N_{k-1} G_k},$$

$$\beta_k^* = \beta_{k-1}^* + N_k G_k \left(T_0 - G_k^T \beta_{k-1}^*\right). \tag{14}$$

### 2.4. Bayesian Optimization.
Given a global optimization problem of an objective function $f$,

$$\mathbf{x}^* = \arg \max f(\mathbf{x}), \tag{15}$$

where $f$ is an expensive black-box function, and $\mathcal{X}$ is the design space of $f(\mathbf{x})$. Besides, $f$ can be evaluated arbitrarily in $\mathcal{X}$. Then, a sequential exploration process is proposed, which, at iteration $n$, location $\mathbf{x}_{n+1}$ is examined at which to evaluate $f$ and observe $y_{n+1}$. After $N$ evaluations, the exploration process terminates, and a final optimal location $\mathbf{x}^*$ is obtained, which is the best optimization result. In the problem of wind speed forecasting, the black-box function $f$ is a wind speed prediction model with hyperparameters $\mathbf{x}$ with a prediction error $y = f(\mathbf{x})$ on a validation dataset. Such $f$ is nonconvex and expensive to evaluate. Bayesian optimization [44] takes advantage of all the optimization function observations to make the sequential exploration process efficient. Bayesian optimization can be described as a sequential model-based optimization method that solves the objective problem. Initially, a probabilistic surrogate model is specified to represent the prior belief on the objective function, and then the posterior belief is calculated as $f$ is evaluated sequentially. The posterior belief represents the belief of $f$ on the observations of the objective function. The typical probabilistic surrogate models include Gaussian process regression, sparse pseudo-input Gaussian process, sparse spectrum Gaussian process, random forest, and gradient boosting decision tree. An acquisition function $\alpha_n : \mathcal{X} \mapsto R$ is used to explore the design space $\mathcal{X}$, incorporating the posterior belief model. It performs exploration and exploitation for the next evaluation of $f$. As a utility function, it measures how optimal a sequence of evaluations is. The acquisition function returns the utility estimate of candidate points for the next evaluation of $f$ and selects $\mathbf{x}_{n+1}$, which produces the maximum utility. The main acquisition functions are the PI (probability of improvement), EI (expected improvement), and UBC (upper confidence bounds). Currently, Bayesian optimization has been demonstrated as a powerful tool for optimal design problems, such as industrial control [45], robotics [46], and chemical experiments [47]. In this paper, a novel Bayesian optimization algorithm, named DART-EI Bayesian optimization is proposed for the wind speed forecasting models. The process of the algorithm is described in Algorithm 2. In the process, the probabilistic surrogate model is dropouts meet multiple additive regression trees (DART) [48], and the acquisition function is the EI. In each iteration of the Bayesian optimization process, the next query point is calculated as follows:

$$\alpha_{\mathrm{EI}}(x) = E\left[\left(y_{\mathrm{optimal}} - m(x)\right)^+\right],$$

$$\mathbf{x}_{n+1} = \arg \max_{x \in \mathcal{X}} \alpha_{\mathrm{EI}n}(\mathbf{x}_n), \tag{16}$$

where $y_{\mathrm{optimal}}$ denotes the best current value, $m(x)$ represents the DART's prediction mean, and $\alpha_{\mathrm{EI}n}(\mathbf{x}_n)$ denotes the EI.

Since the performance of the OSELM model can be impacted by the number of hidden neurons, in this paper, BO is utilized to achieve the optimal performance of OSELM. The objective function of BO is defined as the prediction result of 4-fold cross-validation for OSELM. The input variable of the objective function is the number of hidden neurons, which is a hyperparameter of OSELM. The output variable of the objective function is the mean absolute percent error of 4-fold cross-validation. The objective function is defined as follows:

$$f_{\mathrm{obj}} = \log\left(1 + \mathrm{CV}_4\right), \tag{17}$$

where $\mathrm{CV}_4$ denotes the 4-fold cross-validation loss on the training data set. Besides, the acquisition function is critical, for that it can determine the exploration and exploitation of BO. In this paper, EI is employed as the acquisition function.

### 2.5. Bagging.
Bagging is an efficient ensemble learning algorithm [49]. It can significantly improve the performance of the primary learner. In this paper, Bagging-OSELM is introduced to complete the prediction of high-frequency sublayers of VMD. Initially, the bootstrap sampling method is used to draw two hundred sample data sets $D_1, D_2, \ldots, D_{200}$ from the given training data set $D$. Then, an OSELM $R_i$ is constructed per each data set $D_i$, and the final ensemble model $R$ is built on averaging the prediction values from $R_1, R_2, \ldots, R_{200}$. The detailed Bagging-OSELM algorithm is described as follows (Algorithm 3):

### 2.6. The Performance Evaluation Metrics.
In this paper, the performance of the involved models can be evaluated by the mean absolute error (MAE), the mean absolute percent error (MAPE), and the root mean square error (RMSE). The smaller the evaluation metrics, the better the model performed. The MAE, MAPE, and RMSE are defined as

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left|\widehat{y}_i - y_i\right|,$$

$$\mathrm{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left|\widehat{y}_i - y_i\right|}{y_i} \times 100\%,$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\widehat{y}_i - y_i\right)^2}, \tag{18}$$

```
Bat Algorithm (f)
Input: Target function f (x), x = (x¹, ..., xⁿ)
Initialize the bat population xᵢ with the velocity vᵢ, the pulse frequency fᵢ,
the pulse rates rᵢ and the loudness Aᵢ, i = 1, 2, ..., m.
For each bat bᵢ, do
    Employ equations (5)–(7) to produce new solutions.
    If rand > rᵢ, then
        Choose one candidate solution from the optimal solutions.
    If rand > Aᵢ and f (xᵢ) < f (ẋ), then
        Accept the newly proposed solutions.
        Update rᵢ and Aᵢ by equations (8) and (9).
Return the current best ẋ.
```

ALGORITHM 1: The bat algorithm.

```
Bayesian Optimization (f, Θ, M, N)
Input: Target function f; hyperparameter space Θ; the number of initiation points M; the number of iteration N;
Result: Optimal hyperparameter θ∗
Sample {θᵢ}ᵢ₌₁ᴹ from the hyperparameter space Θ
for i ⟵ 1 to M do yᵢ ⟵ f (θᵢ)
    for j ⟵ 1 to N do
        Fit a surrogate model S on the data set {θᵢ, yᵢ}ᵢ₌₁^(M+j-1), where S denotes a DART model
        Select θ′ ∈ argmax_{θ∈Θ} α(θ, S), where α represents the acquisition function EI
        Update yⱼ ⟵ f (θ′)
Return θ∗ ∈ argmin_{θ∈{θᵢ}ᵢ₌₁ᴺ} yⱼ
```

ALGORITHM 2: The DART-EI Bayesian optimization.

where $\hat{y}_i$ and $y_i$ denote the predicted and observed value at the time $i$, respectively, and $N$ represents the number of data points.

Besides, improved percentage indices $P_{\text{MAE}}$, $P_{\text{MAPE}}$, and $P_{\text{RMSE}}$ are used to compare the performance of two models. The $P_{\text{MAE}}$, $P_{\text{MAPE}}$, and $P_{\text{RMSE}}$ are defined as

$$P_{\text{MAE}} = \left(\frac{\text{MAE1} - \text{MAE2}}{\text{MAE1}}\right) \times 100,$$

$$P_{\text{MAPE}} = \left(\frac{\text{MAPE1} - \text{MAPE2}}{\text{MAPE1}}\right) \times 100, \quad (19)$$

$$P_{\text{RMSE}} = \left(\frac{\text{RMSE1} - \text{RMSE2}}{\text{RMSE1}}\right) \times 100.$$

### 2.7. Pearson's Test.
Pearson's test can evaluate the prediction capability of the involved models. In Pearson's test, the correlation coefficient is calculated to describe the degree of association between the observed data and the predicted data. If the correlation coefficient is 0, then the observed and the predicted values are not correlated. If the coefficient is 1, the observed and the predicted values are 100% correlated. The larger the Pearson correlation coefficient is, the better the model is. Pearson's correlation coefficient can be described as follows:

$$P = \frac{\sum_{i=1}^{N} (Y_i - Y_i)(\hat{Y}_i - \hat{Y}_m)}{\sqrt{\sum_{i=1}^{N} (Y_i - Y_m)^2 \times \sum_{i=1}^{N} (\hat{Y}_i - \hat{Y}_m)^2}}, \quad (20)$$

where $Y_i$ is the actual data, $\hat{Y}_i$ is the forecasting data, $Y_m$ and $\hat{Y}_m$ are the means of the actual data and the forecasting data, respectively, and $N$ denotes the number of data points.

## 3. Case Study

### 3.1. Wind Speed Data Description.
In this paper, two wind speed time-series are used to evaluate the proposed model. These data were collected from the 135-m research towers of the NREL (National Renewable Energy Laboratory) from January 2012 to August 2012. The descriptive statistics of the data are given in Table 1. Each data set contains 1800 points with 10 min interval. Each original data set is divided into a training data set and a test data set. The training data set includes 1–1700 points, and the test data set contains 1701–1800 points. The wind time-series is depicted in Figures 2 and 3, respectively.

### 3.2. Parameter Settings.
In this paper, two kinds of wind speed prediction models are implemented: the single models and the hybrid models. The single models are the GPR model, the LSSVR model, the LSTM model, the OSELM

```
Bagging-OSELM Algorithm
For i = 1 to 200
    Draw train data set D_i from D through the bootstrap method.
    Build a basal OSELM R_i for D_i
Return R(x) = Σ_{i=1}^{200} R_i(x)/200
```

ALGORITHM 3: The Bagging-OSELM.

TABLE 1: Descriptive statistics of the wind speed datasets.

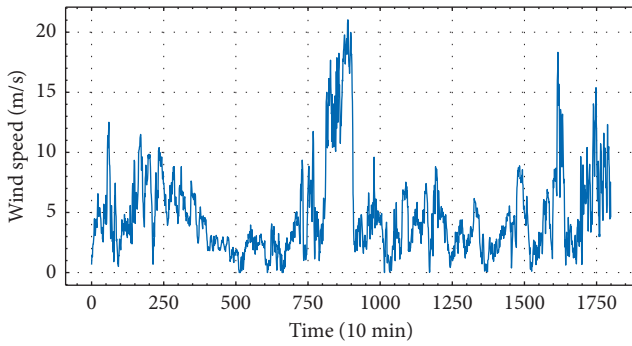| Wind series | Min | Max | Mean | Variance | Skewness | Kurtosis |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 0.0 | 21.03 | 4.78 | 13.52 | 1.74 | 3.48 |
| 2 | 0.0 | 20.50 | 5.00 | 10.99 | 1.10 | 1.04 |



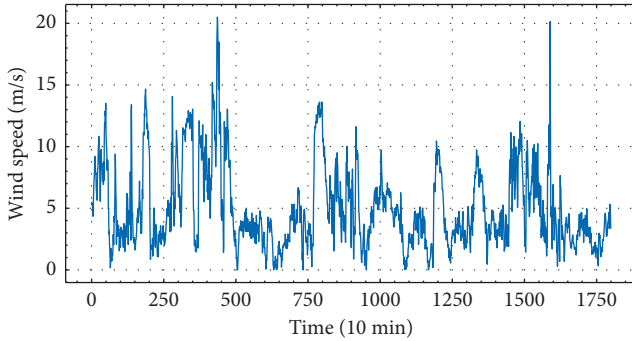FIGURE 2: Original wind speed series 1.



FIGURE 3: Original wind speed series 2.

model, the AdaBoost-OSELM model, the Bagging-OSELM model, the BBA-OSELM model, the BO-OSELM model, and the PSO-OSELM. The hybrid models include the EEMD-BBA-OSELM model, the EMD-BBA-OSELM model, the VMD-BBA-EnsOSELM model, and the VMD-BBA-OSELM model. All the models are developed in the anaconda environment.

In the GPR model, the kernel function is rational quadratic. In LSSVR model, the kernel function is RBF, and gamma is 0.01. In the LSTM model, the number of neurons is 40. In the OSELM models, the number of hidden neurons is 10. In the BBA-based models, the maximum time lag is 20 for selecting relevant input features. In the EMD-BBA-

OSELM model, the number of EMD trials is adopted as 100. In the EEMD-BBA-OSELM model, the number of EEMD trials is adopted as 100, and the standard deviation of Gaussian noise is 0.05 for EEMD. In the VMD-based models (VMD-BBA-OSELM and VMD-BBA-EnsOSELM), the number of modes for VMD decomposition is 10. In the PSO-OSELM model and the VMD-BBA-EnsOSELM model, the number of hidden neurons of OSELM is selected using PSO and BO. The search range is as [10, 200].

*3.3. Experimental Results.* In this section, the forecasting results for wind speed series 1 and 2 are depicted in Figures 4 and 5. The estimation prediction results for wind speed series 1 and 2 are presented in Tables 2 and 3. The improving percentages of the comparison models by the proposed model for wind speed series 1 and 2 are shown in Tables 4 and 5. The results of Pearson's test for wind speed series 1 and 2 are given in Tables 6 and 7.

*3.4. The Comparisons and Analysis.* From the above section, it can be seen that the prediction results for all the wind speed series have similar laws. The comparison and discussion of the prediction results are as follows:

(1) Among the single models, the OSELM model is the most efficient, while the LSTM model is the least efficient. For instance, in series 1, the calculation time of the OSELM model and the LSTM model is 0.02 s and 165.10 s, respectively. In series 2, the calculation time of the OSELM model and the LSTM model is 0.01 s and 140.12 s, respectively.

(2) The ensemble algorithms can improve the prediction accuracy. In series 1, from the OSELM model to the AdaBoost-OSELM model, the MAPE is reduced by 55.73%; from the OSELM model to the Bagging-OSELM model, the MAPE is reduced by 57.80%. In series 2, from the OSELM model to the Bagging-OSELM model, the MAPE is reduced by 64.36%. Besides, Bagging is superior to AdaBoost. For instance, in series 1, from the AdaBoost-OSELM to the Bagging-OSELM model, the MAPE is reduced by 4.67%. In series 2, from the AdaBoost-OSELM to the
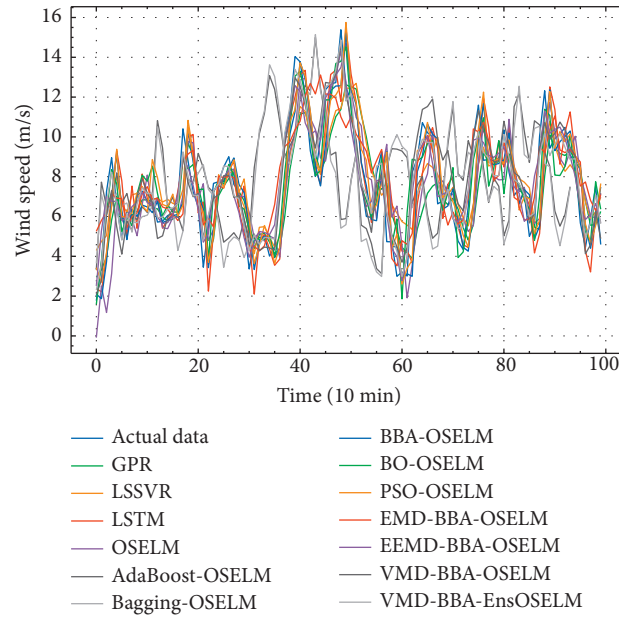
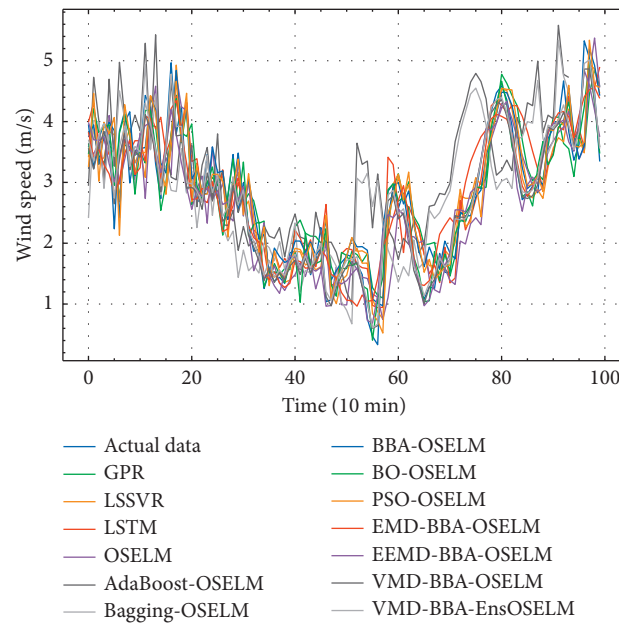FIGURE 4: The forecasting results for wind speed series 1.



FIGURE 5: The forecasting results for wind speed series 2.

Bagging-OSELM model, the MAE is reduced by 23.44%, and the RMSE is reduced by 20.51%

(3) The optimization algorithms can improve the accuracy of the prediction. For instance, in series 1, from the OSELM model to the PSO-OSELM model, the MAPE is reduced by 53.01%, and the RMSE is reduced by 13.01%; from the OSELM to the BO-OSELM model, the MAE is reduced by 40.06%, and the MAPE is reduced by 67.00%. In series 2, from the OSELM model to the PSO-OSELM model, the MAPE is reduced by 66.34%.

(4) The feature selection algorithm of BBA can improve prediction performance. For instance, in series 1, from the OSELM model to the BBA-OSELM model, the MAPE is reduced by 56.67%. In series 2, from the OSELM model to the BBA-OSELM model, the MAPE is reduced by 62.93%.

(5) BO is superior to BBA and PSO. For instance, in series 1, from the BBA-OSELM model to the BO-OSELM model, the MAE is reduced by 30.11%, the MAPE is reduced by 23.83%, and the RMSE is reduced by 29.95%; from the PSO-OSELM model to

TABLE 2: Estimation prediction result for wind speed series 1.

| Model | MAE (m/s) | MAPE (%) | RMSE (m/s) | Time (s) |
|---|---|---|---|---|
| GPR | 1.60 | 22.26 | 2.08 | 42.39 |
| LSSVR | 1.54 | 23.61 | 1.89 | 2.03 |
| LSTM | 1.61 | 23.34 | 1.99 | 165.10 |
| OSELM | 1.64 | 46.58 | 2.11 | 0.02 |
| AdaBoost-OSELM | 1.40 | 20.62 | 1.78 | 5.52 |
| Bagging-OSELM | 1.39 | 19.66 | 1.79 | 1.37 |
| BBA-OSELM | 1.41 | 20.18 | 1.82 | 31.02 |
| BO-OSELM | 0.98 | 15.37 | 1.27 | 102.95 |
| PSO-OSELM | 1.47 | 21.89 | 1.84 | 54.20 |
| EMD-BBA-OSELM | 1.25 | 19.97 | 1.60 | 275.54 |
| EEMD-BBA-OSELM | 0.68 | 10.60 | 0.86 | 219.95 |
| VMD-BBA-OSELM | 0.77 | 11.94 | 0.96 | 219.37 |
| VMD-BBA-EnsOSELM | 0.52 | 8.08 | 0.65 | 455.98 |

TABLE 3: Estimation prediction result for wind speed series 2.

| Model | MAE (m/s) | MAPE (%) | RMSE (m/s) | Time (s) |
|---|---|---|---|---|
| GPR | 0.50 | 25.49 | 0.63 | 46.87 |
| LSSVR | 0.48 | 22.40 | 0.63 | 2.28 |
| LSTM | 0.50 | 23.77 | 0.65 | 140.12 |
| OSELM | 0.54 | 64.95 | 0.70 | 0.01 |
| AdaBoost-OSELM | 0.64 | 33.50 | 0.78 | 2.56 |
| Bagging-OSELM | 0.49 | 23.15 | 0.62 | 1.37 |
| BBA-OSELM | 0.48 | 24.08 | 0.62 | 29.10 |
| BO-OSELM | 0.29 | 14.22 | 0.41 | 98.15 |
| PSO-OSELM | 0.48 | 21.86 | 0.64 | 56.09 |
| EMD-BBA-OSELM | 0.35 | 18.34 | 0.44 | 244.43 |
| EEMD-BBA-OSELM | 0.24 | 11.26 | 0.31 | 243.41 |
| VMD-BBA-OSELM | 0.16 | 7.32 | 0.20 | 336.10 |
| VMD-BBA-EnsOSELM | 0.14 | 6.74 | 0.17 | 536.68 |

TABLE 4: Improving percentages of the comparison models by the proposed model for wind speed series 1.

| Model | $P_{MAE}$ (%) | $P_{MAPE}$ (%) | $P_{RMSE}$ (%) |
|---|---|---|---|
| VMD-BBA-EnsOSELM vs. GPR | 67.69 | 63.72 | 68.72 |
| VMD-BBA-EnsOSELM vs. LSSVR | 66.53 | 65.79 | 65.68 |
| VMD-BBA-EnsOSELM vs. LSTM | 67.93 | 65.40 | 67.37 |
| VMD-BBA-EnsOSELM vs. OSELM | 68.56 | 82.66 | 69.27 |
| VMD-BBA-EnsOSELM vs. AdaBoost-OSELM | 63.15 | 60.84 | 63.49 |
| VMD-BBA-EnsOSELM vs. Bagging-OSELM | 62.81 | 58.92 | 63.67 |
| VMD-BBA-EnsOSELM vs. BBA-OSELM | 63.34 | 59.99 | 64.29 |
| VMD-BBA-EnsOSELM vs. BO-OSELM | 47.55 | 47.47 | 49.02 |
| VMD-BBA-EnsOSELM vs. PSO-OSELM | 64.81 | 63.11 | 64.67 |
| VMD-BBA-EnsOSELM vs. EMD-BBA-OSELM | 58.87 | 59.57 | 59.44 |
| VMD-BBA-EnsOSELM vs. EEMD-BBA-OSELM | 23.86 | 23.84 | 24.77 |
| VMD-BBA-EnsOSELM vs. VMD-BBA-OSELM | 33.36 | 32.39 | 32.35 |

the BO-OSELM model, the MAE is reduced by 32.90% and the RMSE is reduced by 30.70%. In series 2, from the BBA-OSELM model to the BO-OSELM model, the MAPE is reduced by 40.95%; from the

PSO-OSELM model to the BO-OSELM model, the MAE is reduced by 39.58%.

(6) The signal decomposition algorithms can improve the prediction accuracy. For instance, in series 1,

TABLE 5: Improving percentages of the comparison models by the proposed model for wind speed series 2.

| Model | $P_{\mathrm{MAE}}$ (%) | $P_{\mathrm{MAPE}}$ (%) | $P_{\mathrm{RMSE}}$ (%) |
| --- | --- | --- | --- |
| VMD-BBA-EnsOSELM vs. GPR | 72.97 | 73.56 | 73.52 |
| VMD-BBA-EnsOSELM vs. LSSVR | 71.79 | 69.92 | 73.27 |
| VMD-BBA-EnsOSELM vs. LSTM | 72.80 | 71.65 | 74.06 |
| VMD-BBA-EnsOSELM vs. OSELM | 74.85 | 89.62 | 76.04 |
| VMD-BBA-EnsOSELM vs. AdaBoost-OSELM | 78.71 | 79.88 | 78.49 |
| VMD-BBA-EnsOSELM vs. Bagging-OSELM | 72.05 | 70.89 | 72.92 |
| VMD-BBA-EnsOSELM vs. BBA-OSELM | 71.87 | 72.01 | 72.82 |
| VMD-BBA-EnsOSELM vs. BO-OSELM | 53.72 | 52.60 | 59.35 |
| VMD-BBA-EnsOSELM vs. PSO-OSELM | 71.74 | 69.17 | 74.02 |
| VMD-BBA-EnsOSELM vs. EMD-BBA-OSELM | 61.16 | 63.25 | 61.85 |
| VMD-BBA-EnsOSELM vs. EEMD-BBA-OSELM | 42.90 | 40.14 | 46.86 |
| VMD-BBA-EnsOSELM vs. VMD-BBA-OSELM | 17.25 | 7.93 | 18.26 |

TABLE 6: The results of Pearson's test for wind speed series 1.

| Model | Pearson's coefficient |
| --- | --- |
| GPR | 0.6766 |
| OSELM | 0.6948 |
| LSSVR | 0.7223 |
| LSTM | 0.7464 |
| AdaBoost-OSELM | 0.7656 |
| PSO-OSELM | 0.7685 |
| Bagging-OSELM | 0.7741 |
| BBA-OSELM | 0.7806 |
| EMD-BBA-OSELM | 0.8144 |
| BO-OSELM | 0.8851 |
| VMD-BBA-OSELM | 0.9386 |
| EEMD-BBA-OSELM | 0.9540 |
| VMD-BBA-EnsOSELM | 0.9752 |

from the BBA-OSELM model to the EMD-BBA-OSELM model, the MAE is reduced by 10.88%, and the RMSE is reduced by 11.96%. Meanwhile, both EEMD and VMD are superior to EMD. For instance, in series 1, from the EMD-BBA-OSELM model to the EEMD-BBA-OSELM model, the MAE is reduced by 51.86%; from the EMD-BBA-OSELM model to the VMD-BBA-OSELM model, the MAE is reduced by 45.00%, and the RMSE is reduced by 47.21%. In series 2, from the BBA-OSELM model to the EMD-BBA-OSELM model, the RMSE is reduced by 29.03%.

(7) The proposed model performs best among all the involved models. For instance, in series 1, from the OSELM model to the VMD-BBA-EnsOSELM model, the MAE is reduced by 68.56%, the MAPE is reduced by 82.66%, and the RMSE is reduced by 69.27%; from the BBA-OSELM model to the VMD-BBA-EnsOSELM model, the MAE is reduced by 63.34%, the MAPE is reduced by 59.99%, and the RMSE is reduced by 64.29%; from the Bagging-OSELM model to the VMD-BBA-EnsOSELM model, the MAE is reduced by 62.81% and the RMSE is reduced by 63.67%; from the BO-OSELM model to the VMD-BBA-EnsOSELM model, the MAE is reduced by 47.55%, the MAPE is reduced by 47.47%, and the RMSE is reduced by 49.02%; from the VMD-

BBA-OSELM model to the VMD-BBA-EnsOSELM model, the MAE is reduced by 33.36%, and the RMSE is reduced by 32.35%. In series 2, from the OSELM model to the VMD-BBA-EnsOSELM model, the MAE is reduced by 74.07%, and the MAPE is reduced by 89.62%.

(8) Pearson's coefficient of the proposed model is higher than the comparison models in series 1 and series 2, respectively.

### 3.5. The Sensitivity Analysis.
The proposed method involves the number of decomposition modes of VMD, which has to be preconfigured. In this section, several cases are conducted to discuss the sensitivity of the number of modes. The proposed model has performed 1-step predictions for the wind time-series 1 with the various numbers of modes. The forecasting results are shown in Table 8. From Table 8, it is concluded that the prediction errors of the proposed model can be reduced when the number of decomposition modes increases. For instance, when the number of modes grows from 4 to 5, the RMSE index is reduced by 6.62%; when the number of modes grows from 5 to 6, the MAPE index is reduced by 19.90%; when the number of modes grows from 6 to 7, the MAPE index is decreased by 9.19%; when the number of modes grows from 7 to 8, the MAE index is decreased by 19.23%.

TABLE 7: The results of Pearson's test for wind speed series 2.

| Model | Pearson's coefficient |
| --- | --- |
| OSELM | 0.8071 |
| AdaBoost-OSELM | 0.8187 |
| PSO-OSELM | 0.8329 |
| LSTM | 0.8330 |
| GPR | 0.8352 |
| LSSVR | 0.8366 |
| Bagging-OSELM | 0.8483 |
| BBA-OSELM | 0.8494 |
| EMD-BBA-OSELM | 0.9219 |
| BO-OSELM | 0.9315 |
| EEMD-BBA-OSELM | 0.9637 |
| VMD-BBA-OSELM | 0.9872 |
| VMD-BBA-EnsOSELM | 0.9903 |

TABLE 8: Estimation prediction result for various VMD mode numbers.

| Number of modes | MAE (m/s) | MAPE (%) | RMSE (m/s) | Time (s) |
| --- | --- | --- | --- | --- |
| 4 | 1.08 | 16.82 | 1.36 | 148.63 |
| 5 | 1.03 | 16.03 | 1.27 | 223.95 |
| 6 | 0.85 | 12.84 | 1.03 | 241.71 |
| 7 | 0.78 | 11.66 | 0.96 | 253.74 |
| 8 | 0.63 | 9.59 | 0.82 | 287.62 |

## 4. Conclusion

Short-term wind speed forecasting is significant to wind energy development, and it is widely applied to turbine regulation, electricity market clearing, and preload sharing. This paper has presented a novel hybrid forecasting method based on VMD, BBA, BO, Bagging, and OSELM. In the proposed VMD-BBA-EnsOSELM model, VMD is used to decompose the original wind time-series into stationary subseries. BBA is used to complete the feature selection. BO-OSELM and Bagging-OSELM are utilized to complete wind speed prediction. Two experiments are conducted on the NREL datasets to verify the superiority of the proposed method. Twelve involved models are compared with the proposed method, including the GPR model, the LSSVR model, the LSTM model, the OSELM model, the AdaBoost-OSELM model, the Bagging-OSELM model, the BBA-OSELM model, the BO-OSELM model, the PSO-OSELM model, the EMD-BBA-OSELM model, the EEMD-BBA-OSELM model, and the VMD-BBA-OSELM. The experimental results and Pearson's test indicate that (1) BBA is suitable for feature selection; (2) Bagging can be better than AdaBoost for enhancing the prediction capability of OSELM; (3) BO can be superior to PSO for effectively improving the accuracy of a hybrid wind prediction model; (4) the proposed method can achieve the best prediction performance among the involved models. In conclusion, the proposed model fully utilizes the virtues of VMD, BBA, BO, Bagging, and OSELM, and it is suitable for the forecasting of short-term wind speed. Future research directions will focus on enhancing the proposed model for multistep wind speed prediction.

## Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

## References

[1] Z.-X. He, S.-C. Xu, W.-X. Shen et al., "Review of factors affecting China's offshore wind power industry," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 1372–1386, 2016.

[2] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Applied Energy*, vol. 235, pp. 939–953, 2019.

[3] Z. Di, J. Ao, Q. Duan et al., "Improving WRF model turbine-height wind-speed forecasting using a surrogate- based automatic optimization method," *Atmospheric Research*, vol. 226, pp. 1–16, 2019.

[4] Z. Huang and Z. S. Chalabi, "Use of time-series analysis to model and forecast wind speed," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 56, no. 2-3, pp. 311–322, 1995.

[5] J. L. Torres, A. García, M. De Blas, and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Solar Energy*, vol. 79, no. 1, pp. 65–77, 2005.

[6] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renewable Energy*, vol. 34, no. 5, pp. 1388–1393, 2009.

[7] N. Bokde, A. Troncoso, G. Cortés, K. D. Kulat, and F. Martínez-Álvarez, "Pattern sequence similarity based techniques for wind speed forecasting," in *Proceedings of the International Work-Conference on Time Series (ITISE)*, Granada, Spain, October 2017.

[8] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.

[9] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, no. 6, pp. 939–947, 2004.

[10] S. Ding, X. Xu, and R. Nie, "Extreme learning machine and its applications," *Neural Computing and Applications*, vol. 25, no. 3-4, pp. 549–556, 2014.

[11] Y. Zhang, T. Li, G. Na, G. Li, and Y. Li, "Optimized extreme learning machine for power system transient stability prediction using synchrophasors," *Mathematical Problems in Engineering*, vol. 2015, Article ID 529724, 8 pages, 2015.

[12] H. Liu, X. Mi, and Y. Li, "Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM," *Energy Conversion and Management*, vol. 159, pp. 54–64, 2018.

[13] W. Fu, K. Wang, C. Li, and J. Tan, "Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM," *Energy Conversion and Management*, vol. 187, pp. 356–377, 2019.

[14] D. Zhang, X. Peng, K. Pan, and Y. Liu, "A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine," *Energy Conversion and Management*, vol. 180, pp. 338–357, 2019.

[15] 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 Network World*, vol. 28, no. 3, pp. 191–212, 2018.

[16] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, p. 123, 1996.

[17] T. Peng, J. Zhou, C. Zhang, and Y. Zheng, "Multi-step ahead wind speed forecasting using a hybrid model based on two-stage decomposition technique and AdaBoost-extreme learning machine," *Energy Conversion and Management*, vol. 153, pp. 589–602, 2017.

[18] M. Zontul, F. Aydin, G. Dogan, S. Sener, and O. Kaynar, "Wind speed forecasting using REPTree and bagging methods in Kirklareli-Turkey," *Journal of Theoretical and Applied Information Technology*, vol. 56, no. 1, pp. 17–29, 2013.

[19] C. Emeksiz and G. Demir, "An investigation of the effect of meteorological parameters on wind speed estimation using bagging algorithm," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 6, no. 4, pp. 311–321, 2018.

[20] H. Liu, H.-Q. Tian, Y.-F. Li, and L. Zhang, "Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions," *Energy Conversion and Management*, vol. 92, pp. 67–81, 2015.

[21] S.-w. Fei and Y. He, "Wind speed prediction using the hybrid model of wavelet decomposition and artificial bee colony algorithm-based relevance vector machine," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 625–631, 2015.

[22] H. Liu, X. Mi, and Y. Li, "Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network," *Energy Conversion and Management*, vol. 166, pp. 120–131, 2018.

[23] C. Zhang, H. Wei, J. Zhao, T. Liu, T. Zhu, and K. Zhang, "Short-term wind speed forecasting using empirical mode decomposition and feature selection," *Renewable Energy*, vol. 96, pp. 727–737, 2016.

[24] 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.

[25] Y. Zhang, Y. Zhao, and S. Gao, "A novel hybrid model for wind speed prediction based on VMD and neural network considering atmospheric uncertainties," *IEEE Access*, vol. 7, pp. 60322–60332, 2019.

[26] M. Gendeel, Z. Yuxian, and H. Aoqi, "Performance comparison of ANNs model with VMD for short-term wind speed forecasting," *IET Renewable Power Generation*, vol. 12, no. 12, pp. 1424–1430, 2018.

[27] S. Sun, J. Fu, F. Zhu, and N. Xiong, "A compound structure for wind speed forecasting using MKLSSVM with feature selection and parameter optimization," *Mathematical Problems in Engineering*, vol. 2018, Article ID 9287097, 21 pages, 2018.

[28] G. Memarzadeh and F. Keynia, "A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets," *Energy Conversion and Management*, vol. 213, 2020.

[29] N. Huang, E. Xing, G. Cai, Z. Yu, B. Qi, and L. Lin, "Short-term wind speed forecasting based on low redundancy feature selection," *Energies*, vol. 11, no. 7, p. 1638, 2018.

[30] H. Liu, Z. Duan, H. Wu, Y. Li, and S. Dong, "Wind speed forecasting models based on data decomposition, feature selection and group method of data handling network," *Measurement*, vol. 148, Article ID 106971, 2019.

[31] A. K. Naik, V. Kuppili, and D. R. Edla, "Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function," *Soft Computing*, vol. 24, no. 6, pp. 4575–4587, 2020.

[32] X. Xie, X. Qin, Q. Zhou et al., "A novel test-cost-sensitive attribute reduction approach using the binary bat algorithm," *Knowledge-Based Systems*, vol. 186, 2019.

[33] F. Liu, X. Yan, and Y. Lu, "Feature selection for image steganalysis using binary bat algorithm," *IEEE Access*, vol. 8, pp. 4244–4249, 2020.

[34] F. Martínez-Álvarez, G. Asencio-Cortés, J. F. Torres et al., "Coronavirus optimization algorithm: a bioinspired meta-heuristic based on the COVID-19 propagation model," *Big Data*, vol. 8, no. 4, pp. 308–322, 2020.

[35] Y. Wang, S. Wang, and N. Zhang, "A novel wind speed forecasting method based on ensemble empirical mode decomposition and GA-BP neural network," in *Proceedings of the 2013 IEEE Power & Energy Society General Meeting*, Vancouver, Canada, July 2013.

[36] Z. Tian, Y. Ren, and G. Wang, "Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 41, no. 1, pp. 26–46, 2019.

[37] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, "Basic enhancement strategies when using bayesian optimization for

hyperparameter tuning of deep neural networks," *IEEE Access*, vol. 8, pp. 52588–52608, 2020.

[38] P. K. Muhuri and S. K. Biswas, "Bayesian optimization algorithm for multi-objective scheduling of time and precedence constrained tasks in heterogeneous multiprocessor systems," *Applied Soft Computing*, vol. 92, 2020.

[39] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[40] S. Lahmiri, "Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1907–1910, 2017.

[41] M. Zhang, Z. Jiang, and K. Feng, "Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump," *Mechanical Systems and Signal Processing*, vol. 93, pp. 460–493, 2017.

[42] X. Degang and Z. Ping, "Literature survey on research and application of bat algorithm," *Computer Engineering and Applications*, vol. 55, no. 15, pp. 1–12, 2019.

[43] C. Chupong and B. Plangklang, "Comparison study on artificial neural network and online sequential extreme learning machine in regression problem," in *Proceedings of the 2019 7th International Electrical Engineering Congress (iEECON)*, Cha-am, Thailand, 2019.

[44] S. Greenhill, S. Rana, S. Gupta, P. Vellanki, and S. Venkatesh, "Bayesian optimization for adaptive experimental design: a review," *IEEE Access*, vol. 8, pp. 13937–13948, 2020.

[45] M. Neumann-Brosig, A. Marco, D. Schwarzmann, and S. Trimpe, "Data-efficient autotuning with bayesian optimization: an industrial control study," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 3, pp. 730–740, 2020.

[46] K. Junge, J. Hughes, T. G. Thuruthel, and F. Iida, "Improving robotic cooking using batch bayesian optimization," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 760–765, 2020.

[47] A. Seko and S. Ishiwata, "Prediction of perovskite-related structures in ACuO3-x (A = Ca, Sr, Ba, Sc, Y, La) using density functional theory and Bayesian optimization," *Physical Review B*, vol. 101, no. 13, Article ID 134101, 2020.

[48] R. K. Vinayak and R. Gilad-Bachrach, "DART: dropouts meet multiple additive regression trees," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, vol. 38, pp. 489–497, San Diego, CA, USA, May 2015.

[49] M. H. D. M. Ribeiro and L. dos Santos Coelho, "Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series," *Applied Soft Computing*, vol. 86, 2020.