A Novel Decomposition-Optimization Model for Short-Term Wind Speed Forecasting

Jianzhong Zhou [1,2], Na Sun [1,2,*], Benjun Jia [1,2] and Tian Peng [1,2]

[1]   School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; jz.zhou@hust.edu.cn (J.Z.); jbj@hust.edu.cn (B.J.); husthydtopt@126.com (T.P.)

[2]   Hubei Key Laboratory of Digital Valley Science and Technology, Wuhan 430074, China

*   Correspondence: sunna@hust.edu.cn; Tel.: +86-186-2964-9710

check for updates

Abstract: Due to inherent randomness and fluctuation of wind speeds, it is very challenging to develop an effective and practical model to achieve accurate wind speed forecasting, especially over large forecasting horizons. This paper presents a new decomposition-optimization model created by integrating Variational Mode Decomposition (VMD), Backtracking Search Algorithm (BSA), and Regularized Extreme Learning Machine (RELM) to enhance forecasting accuracy. The observed wind speed time series is firstly decomposed by VMD into several relative stable subsequences. Then, an emerging optimization algorithm, BSA, is utilized to search the optimal parameters of the RELM. Subsequently, the well-trained RELM is constructed to do multi-step (1-, 2-, 4-, and 6-step) wind speed forecasting. Experiments have been executed with the proposed method as well as several benchmark models using several datasets from a widely-studied wind farm, Sotavento Galicia in Spain. Additionally, the effects of decomposition and optimization methods on the final forecasting results are analyzed quantitatively, whereby the importance of decomposition technique is emphasized. Results reveal that the proposed VMD-BSA-RELM model achieves significantly better performance than its rivals both on single- and multi-step forecasting with at least 50% average improvement, which indicates it is a powerful tool for short-term wind speed forecasting.

Keywords: wind speed forecasting; hybrid forecasting model; signal decomposition techniques; parameter optimization algorithms

1. Introduction

With the massive consumption of fossil fuel and the increasing pressure of environmental protection, wind energy, one of the most major sustainable and clean energy sources, has been attracting an increasing attention in the last decades due to its remarkable features, such as broad distribution and abundant reserves [1]. Therefore, wind energy is a promising substitute in many parts of the world. As the Global Wind Energy Council (GWEC) have reported, over 54 GW of clean and sustainable wind power has been installed across the global market in 2016, which now contains over 90 countries, including nine with over 10,000 MW installed, and 29 which have now exceeded the 1000 MW mark. Cumulative capacity increased by 12.6% to reach a total of 486.8 GW [1]. However, affected by various factors (e.g., terrain, air pressure, temperature), wind energy is seriously intermittent, random, highly non-linear, and non-stationary, which is not conducive to the large-scale grid-connected operation of wind farms, and can bring a series of fatal problems for the safe and stable operation of power systems. Fortunately, accurate and reliable wind speed forecasting can effectively mitigate the negative impacts of wind energy on the power grid. Thus, many efforts have been done in wind speed forecasting to achieve higher wind energy utilization rates, safe and stable operation of power grids, and thereby gain more economic profits.

At present, various forecasting models have been developed and applied in many fields [2–6]. Weron [3] provided a thorough review of the strengths, weaknesses, and future for the state-of-the-art forecasting methods. Models used in wind speed/power forecasting can be divided into four main types, including physical models, statistical models, machine learning (ML) models, and hybrid models. The physical models are established according to hydrodynamic and thermodynamic equations. They usually require various meteorological and geographic information, such as wind speed, wind direction, temperature, humidity, barometric pressure, air density, elevation, among others. Therefore, the input dimension of the physical models is extremely high and their implementation process are very complex due to the large dimension of inputs. These two features limit the generalization of the physical models in practical engineering applications.

Unlike physical models, statistical models are constructed using relative less historical data through the analysis of the relevance between each point in the observed wind speed series. Most commonly used statistical models are auto regressive (AR) model [7], autoregressive moving average (ARMA) model [8], auto regressive integrated moving average (ARIMA) model [9], and their variants. These models have simple structures, whereas they are often inefficient when handle time series with high-nonlinear and non-stationary characteristics which are two essential features of wind speed series. Therefore, machine learning (ML) models are exploited in this field due to their remarkable abilities of nonlinear learning and generalization abilities. Cincotti et al. [6] has demonstrated that the ARMA-Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) model is inferior to computational intelligence methods. Artificial neural networks (ANNs), the most popular ML models, have been widely exploited over the last decades. Traditional ANNs mainly include multi-layer perceptron (MLP) [6,10], back-propagation neural networks (BPNNs) [11–13], generalized regression neural networks (GRNNs) [13], radial basis function neural networks (RBFNNs) [13], and Elman neural networks (ENNs) [14,15]. Recently, the extreme learning machine (ELM), a new single hidden layer feed-forward network (SLFN), has been developed [16]. Compared with conventional ANNs, the most prominent characteristics of ELM are its simple structure, fast learning rate, and strong generalization ability [16]. Unfortunately, the standard ELM is easy to over-fit and sensitive to outliers, because it only takes the empirical risk minimization principle into account during its implementation process [17–19]. Many researchers have applied their efforts to improving the performance of ELM [17,18]. The most effective way is introducing regularization methods into the basic ELM model to build the regularized ELM (RELM) model. Compared with the basic ELM, the RELM can provide more accurate and stable results, which has been proved by [5,17,18].

With the rapid development of data mining and computational intelligence techniques, a number of hybrid models with signal decomposition approaches and/or optimization algorithms have been proposed/developed. The signal decomposition approaches are able to decompose the raw data into a group of subseries which are smoother and easier to predict. Signal decomposition methods, such as wavelet decomposition (WD) [20,21], empirical mode decomposition (EMD) [22–24], ensemble empirical mode decomposition (EEMD), and variational mode decomposition (VMD) [25,26] are widely used in recent years. Generally, the WD method depends heavily on the determination of the mother wavelet functions, while, EMD has many drawbacks, including lack of an accurate mathematical expression, interpolation method selection, and trapping into mode mixing problems. Although EEMD is capable of solving the mode mixing issues of EMD, it still lacks a mathematical theory, which may reduce its robustness. In contrast, the VMD method can adaptively decompose the raw signal into several modes with specific sparsity properties and is also capable to overcoming the problem of mode mixing [27].

On the other hand, optimization algorithms have become popular in constructing hybrid models by tuning the parameters of ML models to further enhance forecasting accuracy. For example, Ren et al. [11] applied the particle swarm optimization (PSO) algorithm to optimize the parameters of BPNN so as to improve prediction accuracy of wind speed. Similarly, Gao et al. [28] used the firefly algorithm (FA) instead of PSO to adjust the weights and thresholds of the BPNN, and then developed

a new hybrid model. There are more examples of hybrid models based on optimization algorithms in the wind speed/power forecasting, such as BPNN optimized by genetic algorithm (GA) [12], ELM optimized by crisscross optimization algorithm [29], MLP optimized by GA [10], MLP optimized by mind evolutionary algorithm (MEA) [10], SVM optimized by GA [21], least squares support vector machine (LSSVM) optimized by gravitational search algorithm (GSA) [30], and adaptive neuro-fuzzy inference system (ANFIS) optimized by an evolution PSO [31]. Though there are many examples of successful applications for these optimization algorithms, the problems of premature convergence and deficiencies in balancing global search and local mining still exist in these algorithms. Therefore, it is worthwhile to find new efficient algorithms to solve wind speed forecasting problems. Recently, the backtracking search algorithm (BSA), a novel stochastic search algorithm, has been proposed by [32]. Compared with the other stochastic population-based algorithms, BSA needs to set only one control parameter and is easy to implement. Due to its simple structure and easy operation, BSA has been applied to settle various complex nonlinear optimization problems [33–35], and therefore we attempt to use it for solving wind speed forecasting problem in our work.

In this study, a novel decomposition-optimization model is proposed through combining RELM, VMD, and BSA to achieve more accurate and reliable ultra-short-term wind speed forecasting. Firstly, VMD is applied to decompose the original wind speed series into a group of relatively stable subseries to reduce the distractions of the randomness and fluctuations of the original series on the prediction accuracy. Then, RELM optimized by BSA is establish to forecast each subseries. Meanwhile, partial autocorrelation function (PACF) is utilized to determine the optimal input vector. Finally, eventual results can be obtained by the aggregation method. To demonstrate the effectiveness of the proposed model, it has been thoroughly tested on several real wind speed datasets from the Sotavento Galicia (SG) wind farm in Spain. Experimental results demonstrate that by using decomposition and optimization techniques together, the forecasting performance of the proposed VMD-BSA-RELM model is significantly better than that of the basic RELM model. Moreover, the decomposition method VMD plays a more important role in the final improvement of the VMD-BSA-RELM model than the optimization method BSA. This clearly shows how important it is to smooth time series to achieve a desired prediction performance.

The main contributions of this study are listed as follows: (a) we first investigate the ability of the combination of VMD, RELM, and BSA to forecast multi-step short-term wind speed; (b) the proposed model can take full advantages of the signal decomposition approach, machine learning, and optimization algorithm; (c) the positive effects of the decomposition and optimization approaches on the final improvement are quantitatively analyzed.

The rest of the paper is organized as follows: the methods involved in the proposed model including VMD, RELM, and BSA are briefly introduced in Section 2; the framework of the proposed decomposition-optimization model is presented in Section 3; experiments and comprehensive analyses to validate the proposed model are presented in Sections 4 and 5; and Section 6 concludes the paper.

2. Methodology

In this paper, the proposed hybrid model is integrated with three components, variational mode decomposition (VMD), regularized ELM (RELM), and backtracking search algorithm (BSA). So, in this section, separate theories of the VMD-BSA-RELM model will be described in detail.

2.1. Variational Mode Decomposition

Variational mode decomposition (VMD) developed by Dragomiretskiy and Zosso [27] is a novel adaptive and non-recursive signal processing approach. The core of VMD is decomposing a signal $f(t)$ into a series of modes denoted as u_k with specific sparsity characteristic [27]. The sparsity of each mode is called its bandwidth in the spectral domain, which can be estimated using the following steps: (1) Employ the Hilbert transform to each mode u_k to produce a unilateral frequency spectrum, (2) transform frequency spectrum of each mode to baseband regions by means of an exponential

adjusted to the respective estimated frequency, and (3) estimate the bandwidth using the H^1 Gaussian smoothness of the demodulated signal, i.e., L^2-norm of the gradient. Therefore, the process of decomposition is implemented by settling the following optimization problem:

$$
\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \ k = 1, 2, \cdots, K
$$
$$
s.t. \sum_k u_k = f(t) \tag{1}
$$

where u_k and w_k represent the set of all modes and their frequencies, respectively; $f(t)$ denotes the original signal; $\delta(t)$ denotes the Dirac distribution; and * is convolution operator.

Transform the above optimization problem into an unconstrained one by adding a quadratic penalty term and Lagrangian multipliers, as follows:

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

where α denotes the balancing factor of the data-fidelity constraint.

The above unconstrained optimization problem can be solved by means of the ADMM (alternate direction method of multipliers), which can search the saddle point of the augmented Lagrangian in a series of iterative sub-optimizations by updating u_k^{n+1}, ω_k^{n+1}, and λ^{n+1}. u_k^{n+1}, ω_k^{n+1}, and λ^{n+1} are updated by:

$$
\hat{u}_k^{n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}
$$

$$
\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{4}
$$

$$
\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_k^{n+1}(\omega) \right) \tag{5}
$$

where $\hat{u}_k^{n+1}$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$, and $\hat{\lambda}(\omega)$ represent the Fourier transform of u_k^{n+1}, $u_i(t)$, $f(t)$, and $\lambda(t)$, respectively; n denotes the number of iterations; τ is time-step of the dual ascent.

The termination condition of the VMD algorithm is presented as follows:

$$
\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon \tag{6}
$$

where ε is tolerance of convergence criterion. The entire decomposition process for VMD can be described as:

Step 1: Initialize parameters for VMD method including $\hat{u}_k^1$, ω_k^1, and $\hat{\lambda}^1$, and set iteration number $n = 1$.
Step 2: Calculate $\hat{u}_k^{n+1}$ and ω_k^{n+1} using the Equations (3) and (4).
Step 3: Update the Lagrangian multiplier in terms of Equation (5) and then set $n = n + 1$.
Step 4: Repeat the steps 2–4 until meeting the termination condition. Then, the final decomposed modes can be obtained.

2.2. Regularized Extreme Learning Machine

An extreme learning machine (ELM) is a novel single-hidden-layer feed-forward neural network developed by Huang et al. [16]. The significant feature of an ELM is that it randomly generates the input weights and hidden biases, and then determines its output weights directly according to the

Moore-Penrose generalized inverse matrix theory. Suppose there is a given set of training samples (x_t, y_t) with M samples, the output of ELM with L hidden nodes can be estimated by:

$$\hat{y} = \sum_{i=1}^{L} \beta_i g_i(x) = \sum_{i=1}^{L} \beta_i G(w_i \cdot x_t + b_i) \tag{7}$$

where $g_i(x)$ is the activation function of the ith hidden node; w_i is the input weight vector; b_i is the hidden bias and β_i is the output weight connecting the ith hidden node and the output node.

The above equation can be rewritten as:

$$H\beta = Y \tag{8}$$

where $\beta = [\beta_1 \cdots \beta_L]^T$, $Y = [y_1 \cdots y_M]^T$, and H is the hidden layer output matrix defined as :

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_M) \end{bmatrix} = \begin{bmatrix} G(\omega_1 \cdot x_1 + b_1) & \cdots & G(\omega_L \cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ G(\omega_1 \cdot x_M + b_1) & \cdots & G(\omega_L \cdot x_M + b_L) \end{bmatrix}_{M \times L} \tag{9}$$

The output weight can be calculated by means of the least squares method to find the optimal solution of the following equation:

$$\min_{\beta} = \|H\beta - Y\|^2 \tag{10}$$

The optimal solution can be written as:

$$\hat{\beta} = H^{\dagger} Y \tag{11}$$

where $H^{\dagger}$ is the Moore-Penrose generalized inverse matrix of H, which can be calculated by the following orthogonal projection [16]:

$$H^{\dagger} = \left[H^T H\right]^{-1} H^T \tag{12}$$

Due to the numerical instability of the pseudo-inverse, the regularized ELM (RELM) is developed through adding a positive value $1/C$ into the diagonal elements $H^T H$ when calculating the output weights β. Hence, the estimated output weights $\hat{\beta}$ of the RELM can be written as:

$$\hat{\beta} = \left[H^T H + \frac{I}{C}\right]^{-1} H^T Y \tag{13}$$

where I is the identity matrix. More information about RELM can be found in [18].

2.3. Backtracking Search Optimization Algorithm

The backtracking search optimization algorithm (BSA), put forward by Civicioglu [32], is a novel stochastic search algorithm for real-valued numerical optimization problems. In contrast to other population-based evolutionary algorithms, BSA has achieved good performance in both computation speed and computation accuracy. The detailed structure of BSA is described as:

(1) *Initialization*. In this stage, the current population P is randomly generating in the search space by:

$$\begin{aligned} P_{i,j} &= rand(0,1) * (up_j - low_j) + low_j \\ i &= 1, 2, \cdots, N; j = 1, 2, \cdots, D; \end{aligned} \tag{14}$$

where N and D represent the population size and the individual dimensionality, respectively; $rand(0,1)$ is a random generator to provide the number in range (0, 1) uniformly.

(2) *Selection I*. The selection strategy is applied in this process to select the historical population which will guide the search direction in the mutation step. The initial historical population *OldP* is generated by:

$$oldP_{i,j} = rand(0,1) * (up_j - low_j) + low_j$$
$$i = 1, 2, \cdots, N; j = 1, 2, \cdots, D;$$

(15)

At the beginning of each iteration, the *OldP* is updated by:

$$oldP := \begin{cases} P & , a < b \\ permuting(oldP) & , otherwise \end{cases}$$

(16)

where *a* and *b* are two random numbers distributed in the range (0, 1) uniformly; *permuting(oldP)* means that the order of the individuals in *oldP* is randomly updated by a shuffling function.

(3) *Mutation*. In this step, the initial form of the trial population *Mutant* is defined as:

$$Mutant = P + F \cdot (oldP - P)$$
$$F = 3 \cdot rndn, \ rndn \sim N(0,1)$$

(17)

where $(oldP - P)$ is the search-direction matrix; *F* is the mutation factor, which controls the amplitude of $(oldP - P)$.

(4) **Crossover**. In this step, the final form of the trial population *T* is generated. The crossover operator contains a two-stage process. In the first step, a binary integer-valued matrix *map* of size $N \times D$ is generated by:

$$\begin{cases} map(i, u(1 : mixrate * rand * D)) = 0, & when \ c < d | c, d \sim U(0,1) \\ map(i, randi(D)) = 0, & otherwise \end{cases}$$

(18)

where $u = permuting(D)$ represents that the order $1, 2, \cdots, D$ is changed by a random shuffle function; *mixrate* is the only control parameter in BSA (called the mix rate parameter), which controls the number of the individuals that will mutate in a trial.

In the second step, the trial population *T* is updated by:

$$T_{ij} = P_{ij}, when \ map_{ij} = 1; i = 1, 2, \cdots, N; j = 1, 2, \cdots, D$$

(19)

Note that, several individuals of the final trial population *T* may exceed the permissible search space, hence boundary control strategy is quite necessary. The boundary control strategy is:

$$T_{ij} = rnd \cdot (up_j - low_j) + low_j, \ T_{ij} < low_j \ or \ T_{ij} < up_j$$

(20)

(5) *Selection II*. A greedy selection is applied in this stage to update the population *P*, trial individuals with better fitness value then are reserved. Steps 2–5 are repeated until the terminal condition is reached.

3. The Proposed Decomposition-Optimization Model

The decomposition-optimization model developed in this study consists of variational mode decomposition (VMD), regularized ELM (RELM), and backtracking search algorithm (BSA). The proposed decomposition-optimization model is shorted as VMD-BSA-RELM. In the VMD-BSA-RELM model, VMD is first used to smooth the wind speed data for preprocessing. RELM is adopted as a predictor. Meanwhile, partial autocorrelation function (PACF) is executed to choose the suitable input vector and BSA is applied to optimize the input weights and hidden thresholds of the RELM model. The detailed procedures of the proposed hybrid VMD-BSA-RELM model are shown in Figure 1.

Due to the multi-step wind speed forecasting can provide more useful information for decision makers, so the proposed VMD-BSA-RELM model is executed for multi-step wind speed forecasting. The input-output combinations for different forecasting horizons are shown as:

$$\hat{y}_{t+h} = f(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) \tag{21}$$

where *h* is forecasting horizon; *d* is the suitable lag time which is determined by the PACF.
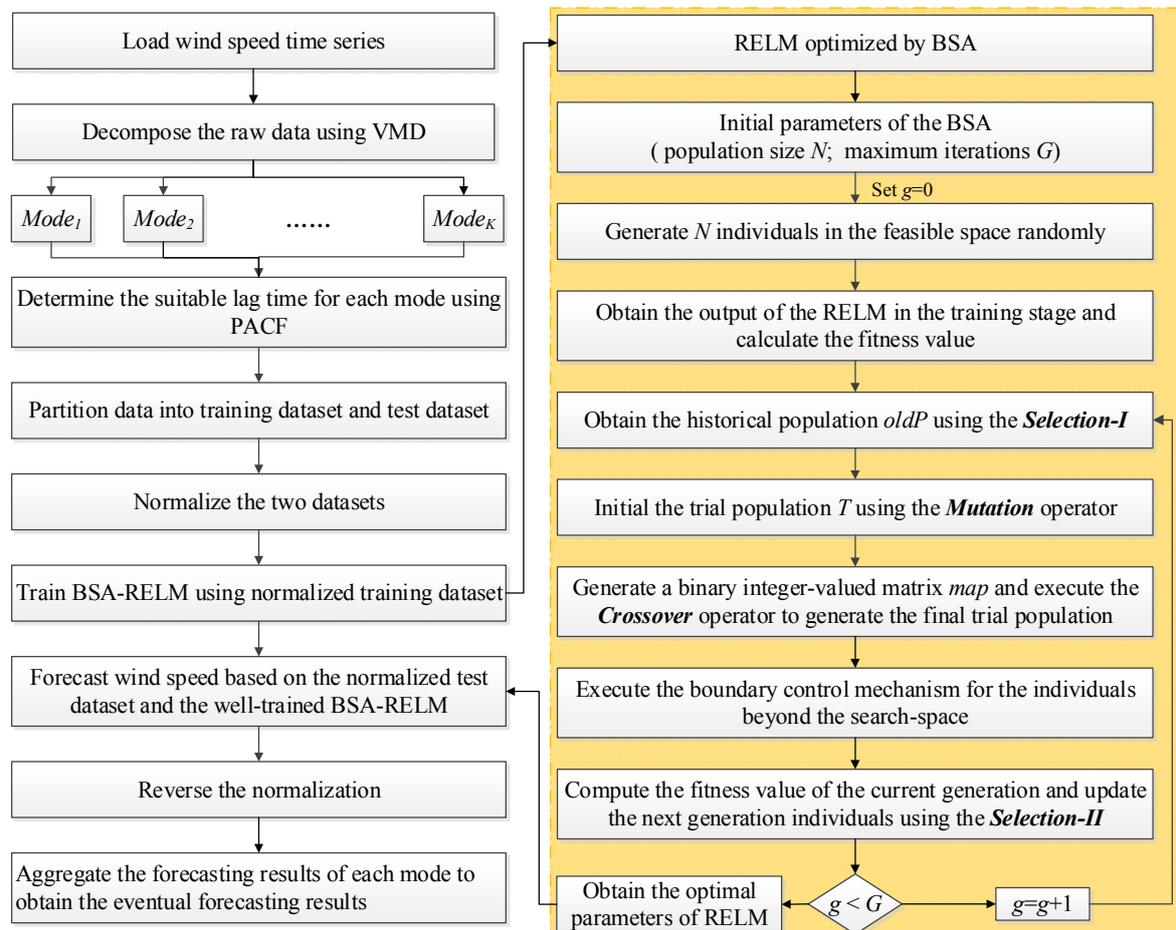

Figure 1. The detailed procedures of the proposed VMD-BSA-RELM model.

4. Experimental Design

4.1. Data Collection

In this study, historical wind speed data were collected from the Sotavento Galicia (SG) wind farm (original wind speed data from the SG wind farm can be found at: http://sotaventogalicia.com/en/real-time-data/historical.). The SG wind farm is located in Galicia, in northwest Spain, with latitude/longitude of 43.354377° N and 7.881213° W. Considering the influence of seasonal factors on forecasting accuracy, four datasets, A, B, C, and D, from different seasons were selected to verify the effectiveness of the proposed VMD-BSA-RELM method. Time periods of the four datasets are 15–21 January, 17–23 April, 13–19 July, and 3–9 October, respectively. Each dataset includes 1008 points with 10 min interval. Based on our test results and [36–39], in each dataset, the first 75% data are selected as training samples to build the prediction model while the remaining 25% are utilized to test. The proposed model is applied to obtain 1-step, 2-step, 4-step, and 6-step (1 h) ahead wind speed

forecasting. The raw wind speed data are shown in Figure 2, which indicates non-stationary and nonlinear features of wind speed series.

The statistical information including average (Ave.) value, maximum (Max.) value, minimum (Min.) value, standard deviation (Std.), the coefficient of variation (C_v) and the skewness coefficient (C_s) of the four datasets are listed in Table 1. The standard deviations of all datasets are all above 1.49 (m/s), and the maximum/minimum values of Datasets A-D are 15.91/0.35 (m/s), 19.13/0.64 (m/s), 9.94/0.35 (m/s), and 9.08/0.74 (m/s). These results also indicate the non-stationary and nonlinear features of the original wind speed series.
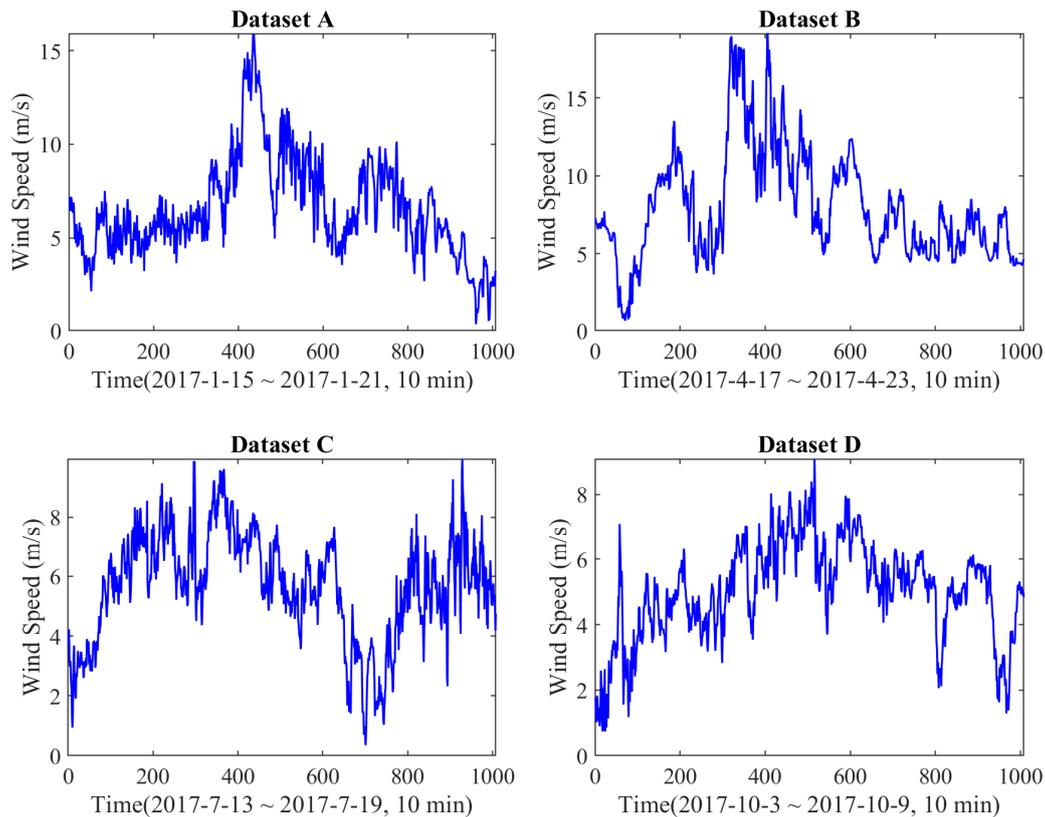

Figure 2. Wind speed on different time periods.

Table 1. Statistical information for the four datasets.

Seasons	Dataset	Ave. (m/s)	Std. (m/s)	C_v.	C_s.	Max. (m/s)	Min. (m/s)	Numbers
Winter	A	6.52	2.64	0.40	0.85	15.91	0.35	1008
Spring	B	7.96	3.53	0.44	0.85	19.13	0.64	1008
Summer	C	5.68	1.80	0.32	−0.43	9.94	0.35	1008
Autumn	D	5.12	1.49	0.29	−0.46	9.08	0.74	1008

4.2. Data Decomposition and Parameters Settings

VMD is executed to decompose the raw wind speed series into several relatively stable modes to make them easy to be predicted. Before the implementation of the decomposition using VMD, the number of modes *K* needs to be preset. In this study, number of modes for each wind speed series is searched in the range [3,14], respectively. Then, the suitable number of modes is determined by the center pulsation of the decomposed modes [19]. After that, each mode will be forecasted by the RELM optimized by BSA (BSA-RELM for short). The input vector of the BSA-RELM is determined by the partial autocorrelation function (PACF). Take Dataset A as an example, the subseries generated by VMD are shown in Figure 3 and the PACF values with 95% confidence interval are presented

in Figure 4. According to the partial autocorrelograms in Figure 4, the lagged variable with PACF value over the confidence interval will be chosen to form the input vector of forecasting model. The population size and the maximum iterations of BSA are set to 50 and 100, respectively. The input selection approach PACF is exploited for all forecasting models involved in this study to guarantee fair and effective comparisons.
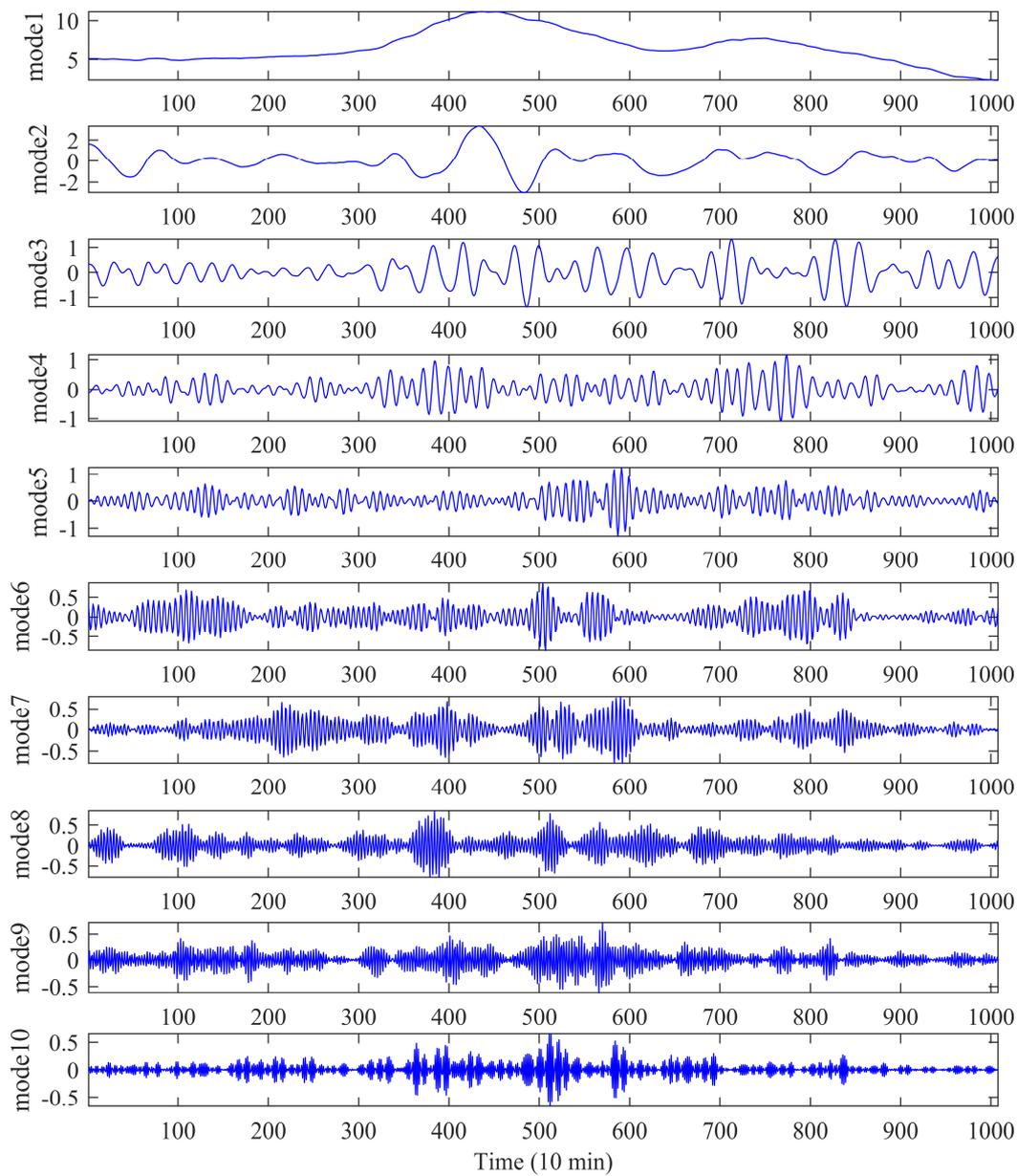

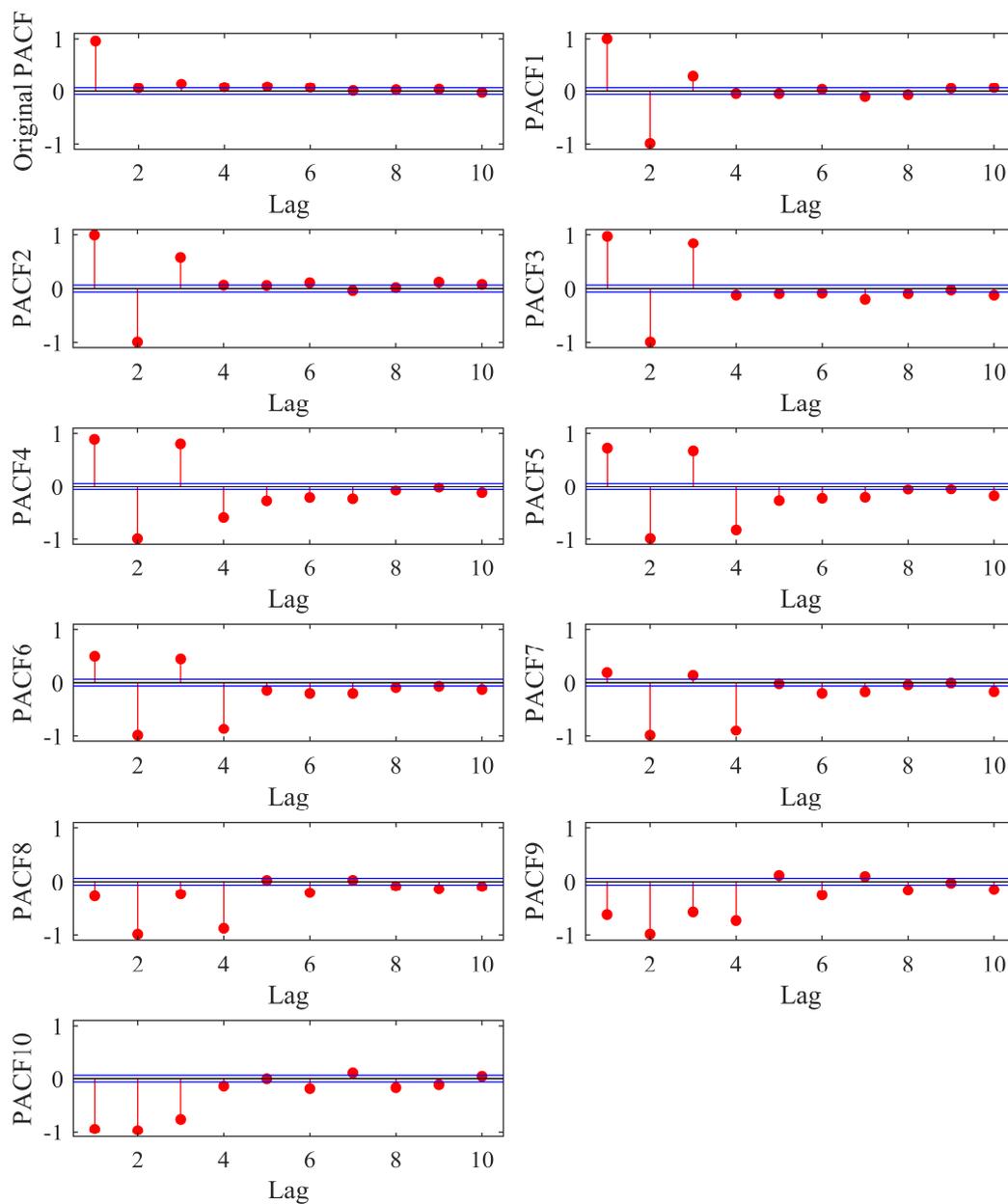**Figure 3.** Decomposed subseries generated by VMD for Dataset A.

Figure 4. PACF values of the original series and the subseries for Dataset A.

4.3. Evaluation Indices

To evaluate forecasting performance of all forecasting models, three commonly used error evaluation metrics including mean absolute error (*MAE*), mean absolute percent error (*MAPE*), and root mean square error (*RMSE*) are used in the study. They can be calculated by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{22}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{23}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{24}$$

where y_i and $\hat{y}_i$ are the ith observed and predicted wind speed, respectively; N is the number of samples.

To clearly view the improvement of a specific model, improved percentage metrics of *RMSE*, *MAE*, and *MAPE* including P_{MAE}, P_{RMSE}, and P_{MAPE} are calculated to exhibit the relative improvement degree between two different models denoted as Model 1 and Model 2. P_{MAE}, P_{RMSE}, and P_{MAPE} of Model 2 relative to Mode 1 can be defined as:

$$P_{MAE} = 100 \times (MAE_1 - MAE_2)/MAE_1 \tag{25}$$

$$P_{RMSE} = 100 \times (RMSE_1 - RMSE_2)/RMSE_1 \tag{26}$$

$$P_{MAPE} = 100 \times (MAPE_1 - MAPE_2)/MAPE_1 \tag{27}$$

5. Results and Discussions

Several experimental results are presented in this section to demonstrate the efficiency and applicability of the proposed decomposition-optimization model (VMD-BSA-RELM). These experiments are grouped into three subsections: one-step forecasting results, multi-step forecasting results, and Diebold-Mariano tests and computational time.

5.1. One-Step Forecasting Results

This part focuses on presenting the one-step forecasting performance of the proposed VMD-BSA-RELM forecasting model using four datasets from different seasons. ARIMA, RBF, GRNN, RELM, VMD-RELM, and BSA-RELM are used as comparison models. Akaike's Information Criteria (AIC), which has widely used in model selections [40,41], is used to determine the appropriate parameters of ARIMA. *RMSE, MAE,* and *MAPE* values provided by these seven forecasting models on the testing data for all datasets are exhibited in Table 2, where the model with the lowest evaluation indices values are highlighted in green. It can be seen that as for the three single neural network models (RELM, GRNN, and RBF), RELM has the best performance and GRNN has the worst performance for all datasets from different seasons. Meanwhile, forecasting results of ARIMA are closer to that of RELM in some cases. Further, comparisons of RELM and VMD-RELM, RELM and BSA-RELM, RELM and VMD-BSA-RELM suggest hybrid models outperform single model in all cases. This can be directly and clearly seen in Table 3 and Figure 5. It is clear that both VMD and BSA have positive effects on improving forecasting accuracy, while BSA has less contribution than VMD. Figure 5 visually indicates that the decomposition-optimization method can gain remarkable improvement of forecasting accuracy compared with hybrid models based on either signal decomposition approach or optimization algorithm. This clearly shows how important it is to incorporate VMD, BSA, and RELM to achieve a desired prediction performance. Concretely, the average improved percentages of *RMSE, MAE,* and *MAPE* between the VMD-BSA-RELM model and the single RELM model are 65.60%, 65.88%, and 66.21%, respectively, indicating a remarkable improvement. Figure 5 also shows that P_{RMSE}, P_{MAE}, and P_{MAPE} values between VMD-BSA-RELM and VMD-RELM are greater than those between VMD-RELM and RELM. Similarly, P_{RMSE}, P_{MAE}, and P_{MAPE} values between VMD-BSA-RELM and BSA-RELM are greater than those between BSA-RELM and RELM. These results emphasize the importance of the signal decomposition approach VMD, and the proposed models can take full advantages of both decomposition and optimization techniques. Predicted and observed curves as well as forecasting errors of all forecasting models are shown in Figure 6, where the predicted curves of the VMD-BSA-RELM model are close to real curves and its forecasting errors are evenly distributed around zero with a tiny range.

Table 2. One-step forecasting results for different models.

Model	Winter (Dataset A)			Spring (Dataset B)		
	RMSE (m/s)	*MAE* (m/s)	*MAPE* (%)	*RMSE* (m/s)	*MAE* (m/s)	*MAPE* (%)
ARIMA	0.9089	0.6718	22.63	0.4943	0.3541	5.97
GRNN	1.1120	0.8674	38.15	0.5480	0.4080	7.01
RBF	0.7627	0.5808	24.09	0.5251	0.3923	6.71
RELM	0.7562	0.5515	17.97	0.4825	0.3531	6.02
VMD-RELM	0.4597	0.3583	15.30	0.2461	0.1817	3.05
BSA-RELM	0.6397	0.4866	17.16	0.4708	0.3279	5.47
VMD-BSA-RELM [a]	0.1821	0.1342	3.80	0.1940	0.1351	2.29
Model	Summer (Dataset C)			Autumn (Dataset D)		
	RMSE (m/s)	*MAE* (m/s)	*MAPE* (%)	*RMSE* (m/s)	*MAE* (m/s)	*MAPE* (%)
ARIMA	0.8448	0.6508	11.74	0.3663	0.2837	7.86
GRNN	0.8563	0.6653	12.10	0.7688	0.5976	19.03
RBF	0.8307	0.6481	11.73	0.6297	0.3884	13.68
RELM	0.8341	0.6469	11.67	0.3673	0.2767	7.48
VMD-RELM	0.5012	0.3829	7.09	0.1588	0.1128	3.42
BSA-RELM	0.8275	0.6533	11.79	0.3060	0.2231	6.50
VMD-BSA-RELM [a]	0.3204	0.2546	4.52	0.1291	0.0952	2.78

[a] The model with the lowest RMSE, MAE, and MAPE values are marked in green.
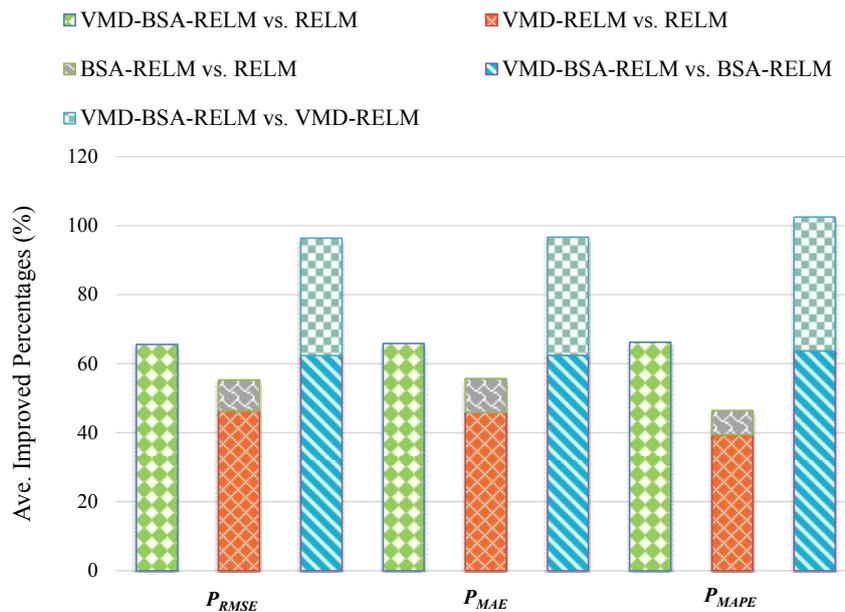

Figure 5. Average improved percentage of different error indices for all four datasets under one-step forecasting.
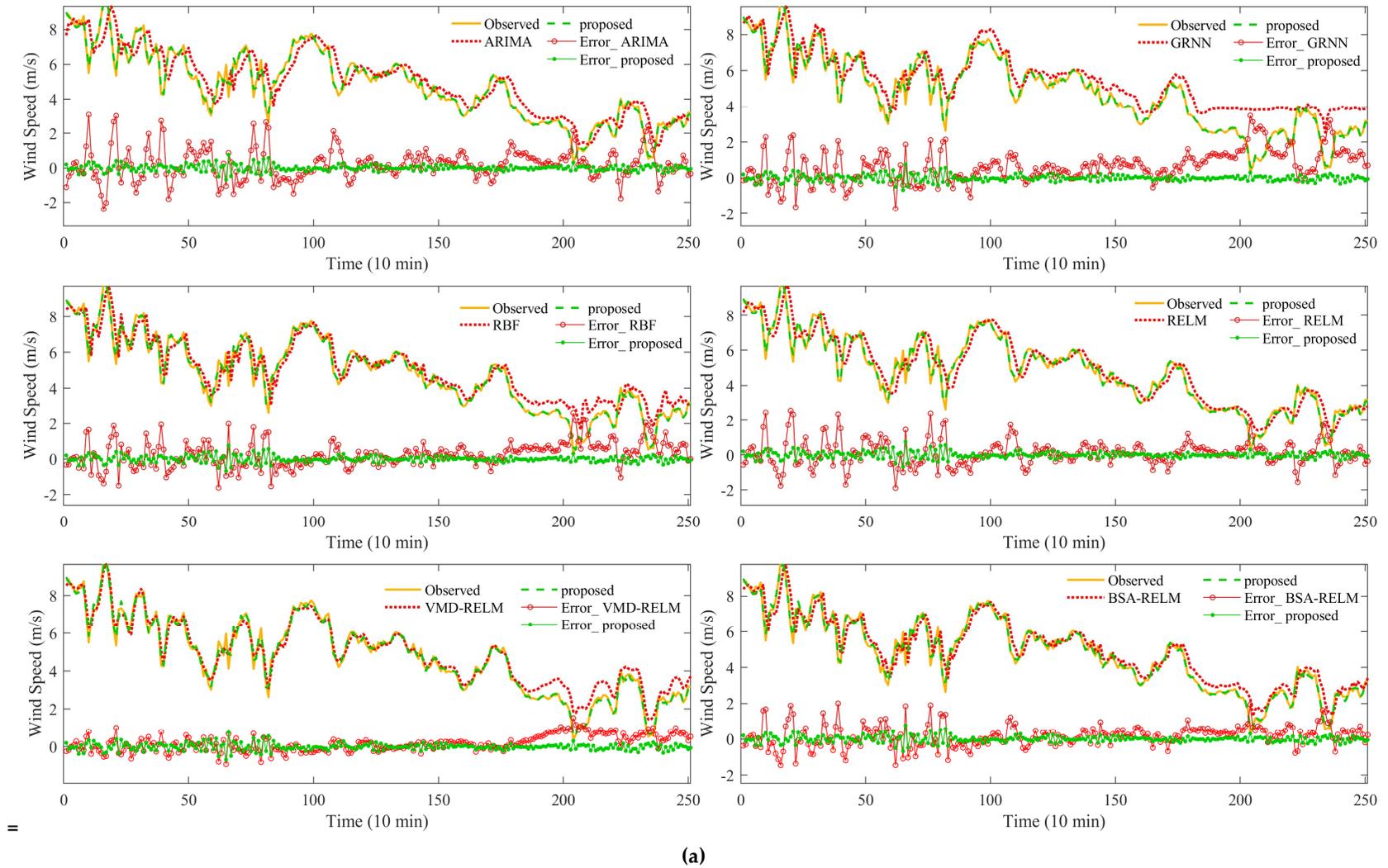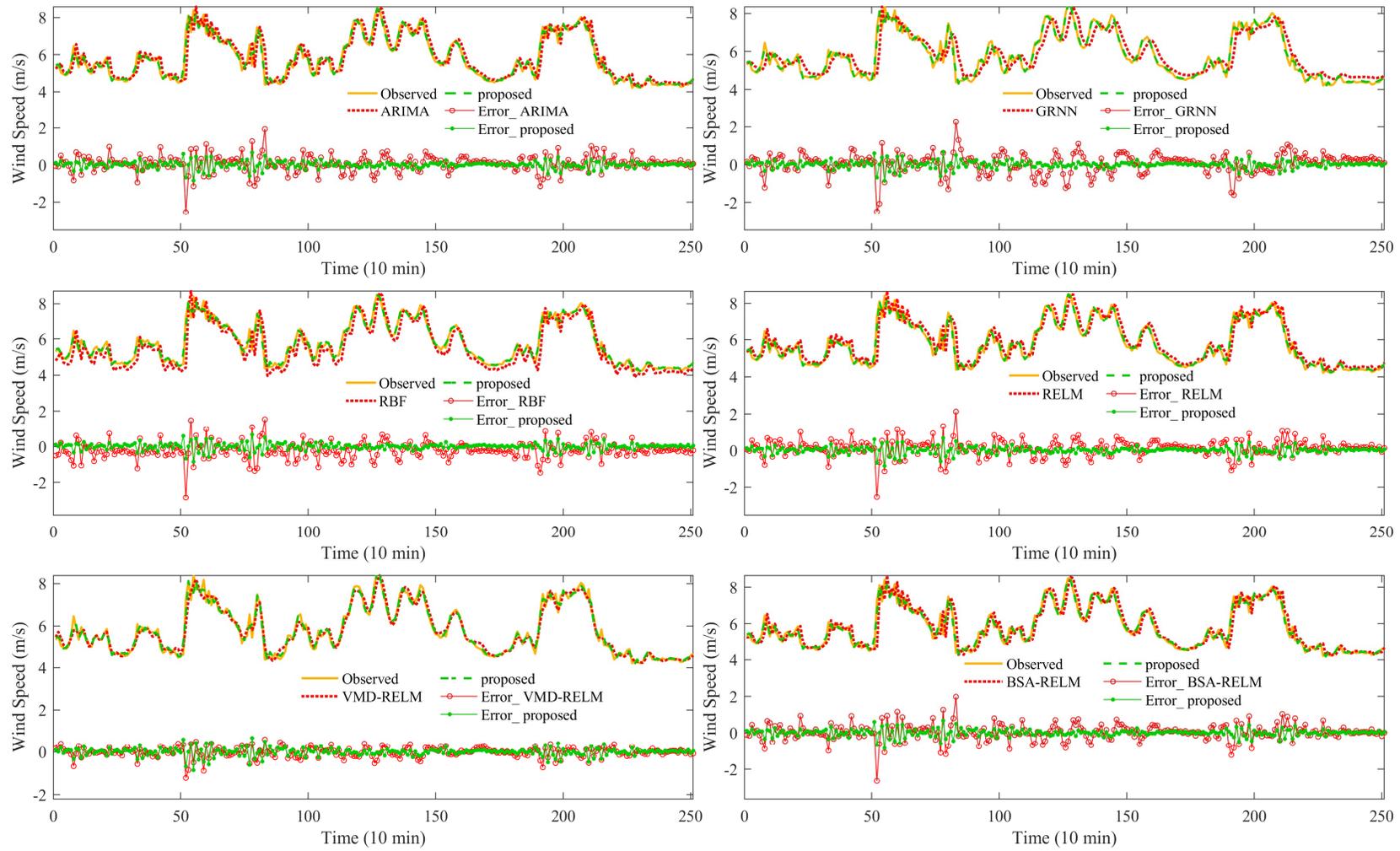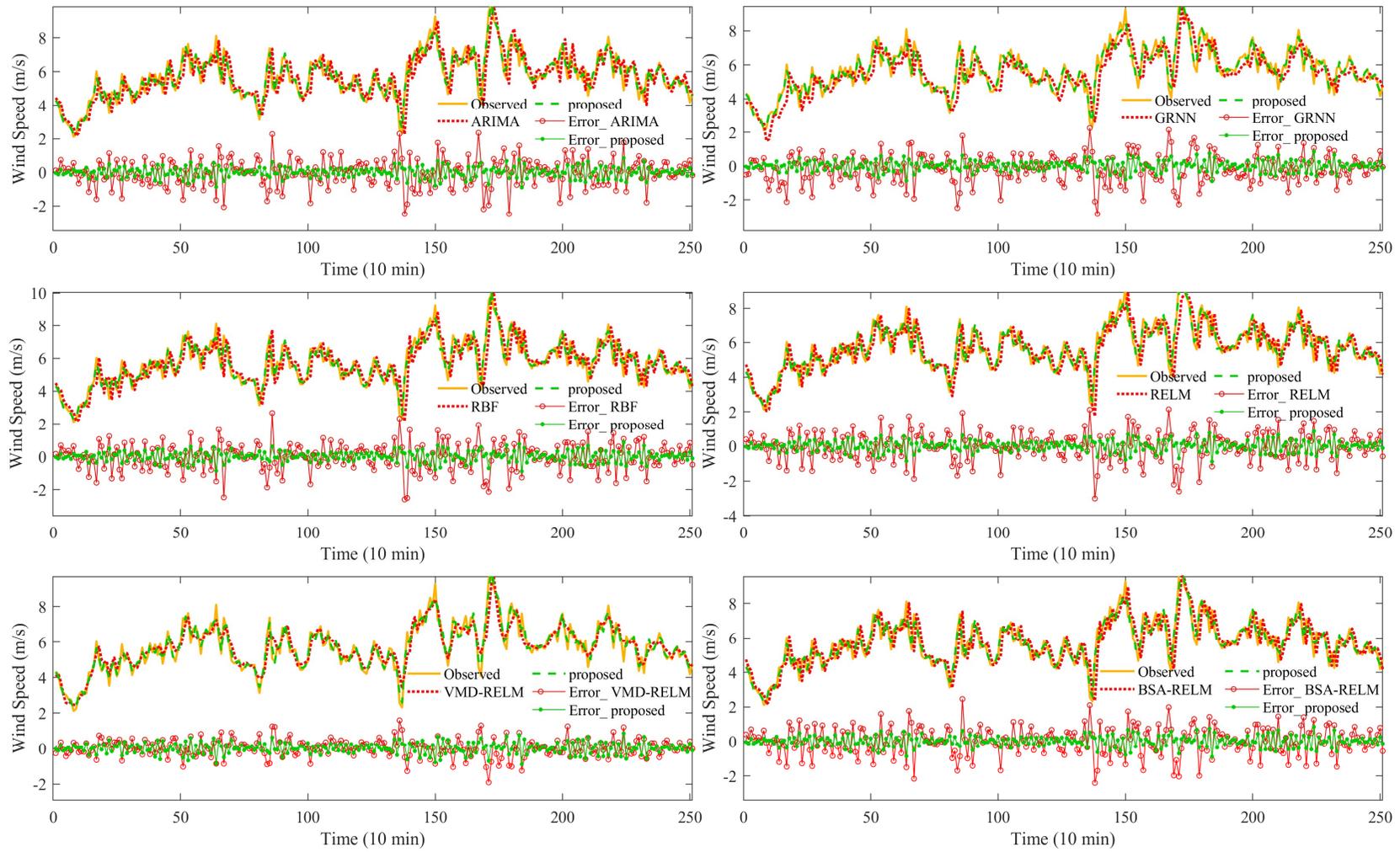
(a)

Figure 6. *Cont.*

(b)

Figure 6. *Cont.*

(c)

Figure 6. *Cont.*

Figure 6. (**a**) One-step forecasting results for seven forecasting models for Dataset A. (**b**) One-step forecasting results for seven forecasting models for Dataset B. (**c**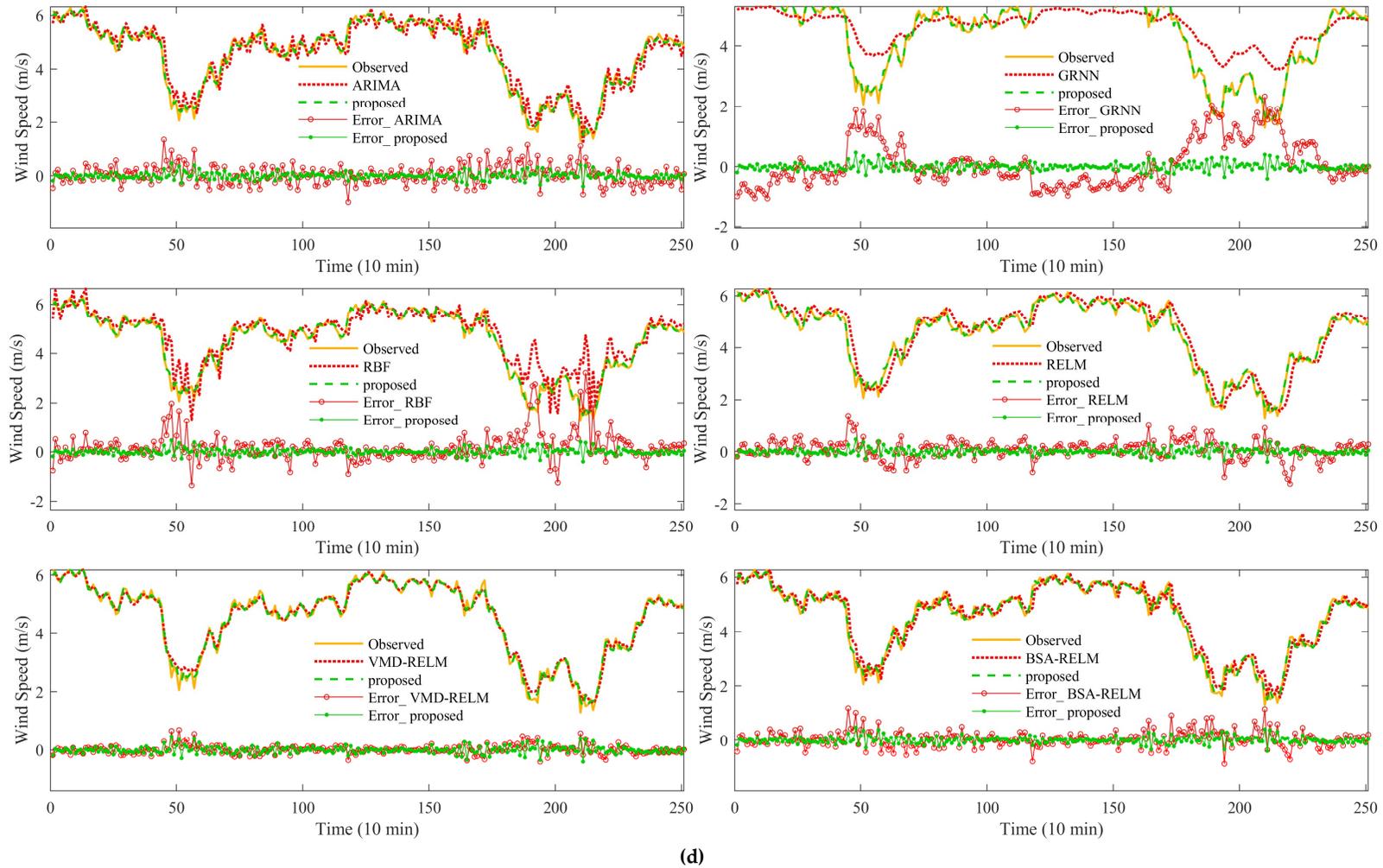) One-step forecasting results for seven forecasting models for Dataset C. (**d**) One-step forecasting results for seven forecasting models for Dataset D.

Table 3. Percentage of improvement of different error indices under one-step forecasting.

Datasets	Forecasting Models	Improved Percentage of Different Error Indices		
		P_{RMSE} (%)	P_{MAE} (%)	P_{MAPE} (%)
Dataset A (Winter)	VMD-RELM vs. RELM	39.2	35.04	14.88
	BSA-RELM vs. RELM	15.41	11.77	4.5
	proposed vs. RELM	75.92	75.67	78.85
	proposed vs. VMD-RELM	60.39	62.55	75.16
	proposed vs. BSA-RELM	71.53	72.43	77.86
Dataset B (Spring)	VMD-RELM vs. RELM	48.99	48.54	49.29
	BSA-RELM vs. RELM	2.43	7.13	9.15
	proposed vs. RELM	59.79	61.75	61.97
	proposed vs. VMD-RELM	21.16	25.66	25
	proposed vs. BSA-RELM	58.79	58.81	58.14
Dataset C (Summer)	VMD-RELM vs. RELM	40.3	40.66	39.08
	BSA-RELM vs. RELM	1.42	1.25	1.3
	proposed vs. RELM	61.83	60.55	61.12
	proposed vs. VMD-RELM	36.06	33.52	36.18
	proposed vs. BSA-RELM	61.28	61.03	61.62
Dataset D (Autumn)	VMD-RELM vs. RELM	56.76	59.21	54.27
	BSA-RELM vs. RELM	16.69	19.35	13.17
	proposed vs. RELM	64.86	65.57	62.9
	proposed vs. VMD-RELM	18.74	15.59	18.87
	proposed vs. BSA-RELM	57.82	57.31	57.27
Average	VMD-RELM vs. RELM	46.31	45.86	39.38
	BSA-RELM vs. RELM	8.99	9.88	7.03
	proposed vs. RELM	65.60	65.88	66.21
	proposed vs. VMD-RELM	34.09	34.33	38.80
	proposed vs. BSA-RELM	62.36	62.40	63.72

5.2. Multi-Step Forecasting Results

This section is devoted to illustrate the efficacy of the proposed model on multi-step wind speed forecasting. The RELM, VMD-RELM, and BSA-RELM which perform better among all competitors are performed as benchmark models in this experiment. Table 4 displays the forecasting performance for different seasons by these four models in 2-step, 4-step, and 6-step forecasting in terms of *RMSE*, *MAE*, and *MAPE* values, where the lowest values among diverse models are emphasized in green. It can be seen that although the forecasting performances deteriorate as the length of the forecasting horizons increase, the proposed model can always outperform than other forecasting models in all cases and horizons, followed by the VMD-RELM model, last the RELM. For instance, for the Dataset A (Winter), the *RMSE* value of the proposed model (VMD-BSA-RELM) in 2-step, 4-step, and 6-step are 0.2462 m/s, 0.4286 m/s, and 0.7120 m/s, respectively, which are better than these of 0.604 m/s, 0.7544 m/s and 1.0633 m/s for VMD-RELM. Moreover, the performance of BSA-RELM is closer to or even worse than that of the RELM along with the increase of the forecasting steps. More concretely, for the Dataset C, the *RMSE* and *MAE* values of the BSA-RELM are 1.2911 m/s and 1.0298 m/s for 6-step forecasting, which are slightly worse than these of 1.2893 m/s and 1.0286 m/s for the single RELM model.

Table 4. Multi-step forecasting results for different models.

Datasets	Models	2-Step			4-Step			6-Step		
		RMSE (m/s)	*MAE* (m/s)	*MAPE*	*RMSE* (m/s)	*MAE* (m/s)	*MAPE*	*RMSE* (m/s)	*MAE* (m/s)	*MAPE*
A	RELM	0.9646	0.7208	0.2466	1.1927	0.9507	0.3319	1.3488	1.0901	0.3947
	VMD-RELM	0.6040	0.4746	0.1943	0.7544	0.6105	0.2449	1.0663	0.8491	0.3378
	BSA-RELM	0.9379	0.7083	0.2564	1.1868	0.9466	0.3385	1.3411	1.0854	0.4043
	Proposed [a]	0.2462	0.1833	0.0563	0.4286	0.3172	0.0909	0.7120	0.5422	0.1703
B	RELM	0.7569	0.5361	0.0901	0.9993	0.7575	0.1301	1.1419	0.8928	0.1570
	VMD-RELM	0.3230	0.2441	0.0414	0.4664	0.3543	0.0600	0.5915	0.4446	0.0746
	BSA-RELM	0.6737	0.4665	0.0794	0.9389	0.7000	0.1203	1.1039	0.8708	0.1537
	Proposed	0.2036	0.1482	0.0250	0.4174	0.3148	0.0540	0.4239	0.3228	0.0550
C	RELM	1.0677	0.8089	0.1528	1.2245	0.9533	0.1793	1.2893	1.0286	0.1915
	VMD-RELM	0.6050	0.4562	0.0838	0.8118	0.6347	0.1177	0.8627	0.6578	0.1228
	BSA-RELM	1.0538	0.8036	0.1499	1.2227	0.9342	0.1761	*1.2911*	*1.0298*	0.1899
	Proposed	0.3809	0.2990	0.0542	0.6202	0.4626	0.0862	0.6618	0.4959	0.0919
D	RELM	0.4861	0.3660	0.1032	0.6650	0.4800	0.1454	0.8041	0.5616	0.1778
	VMD-RELM	0.1935	0.1434	0.0425	0.3064	0.2287	0.0655	0.4344	0.3232	0.0927
	BSA-RELM	0.4345	0.3241	0.0927	0.6336	0.4617	0.1394	0.7725	0.5434	0.1706
	Proposed	0.1377	0.0995	0.0296	0.1957	0.1496	0.0432	0.3301	0.2528	0.0717
Average	RELM	0.8188	0.6080	0.1482	1.0204	0.7854	0.1966	1.1460	0.8933	0.2302
	VMD-RELM	0.4314	0.3296	0.0905	0.5848	0.4570	0.1220	0.7387	0.5687	0.1570
	BSA-RELM	0.7750	0.5756	0.1446	0.9955	0.7606	0.1936	1.1271	0.8823	0.2296
	Proposed	0.2421	0.1825	0.0413	0.4155	0.3110	0.0686	0.5319	0.4034	0.0972

[a] The model with the lowest *RMSE*, *MAE*, and *MAPE* values are marked in green.

The 4-step and 6-step wind speed forecasting results as well as forecasting errors for different models are presented in Figures 7–10, where the superiority of the VMD-BSA-RELM model is confirmed. In these figures, the proposed VMD-BSA-RELM model always provides the smallest forecasting error variation ranges than other models and can accurately capture the variation trend of real wind speed, even for the 6-step forecasting. Additionally, forecasting errors increase along with the growth of horizons, specifically in peak and valley parts. Overall, the proposed model can maximize the advantages of the VMD and BSA methods to produce highly accurate results in multi-step forecasting, which is consistent with the conclusion drawn from one-step forecasting results.
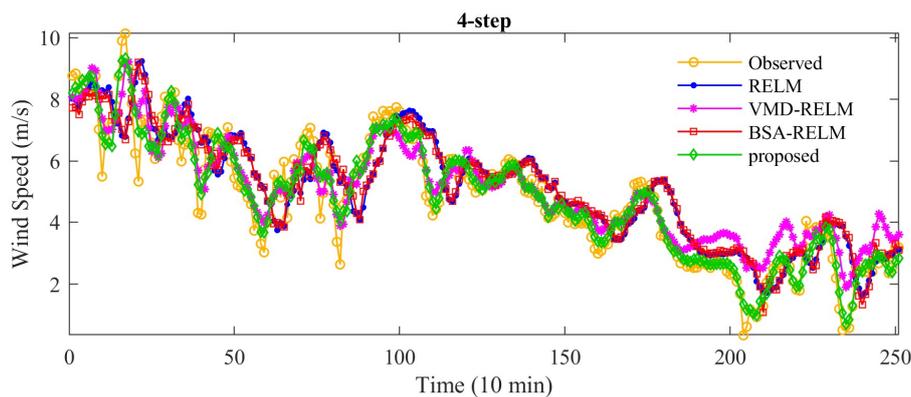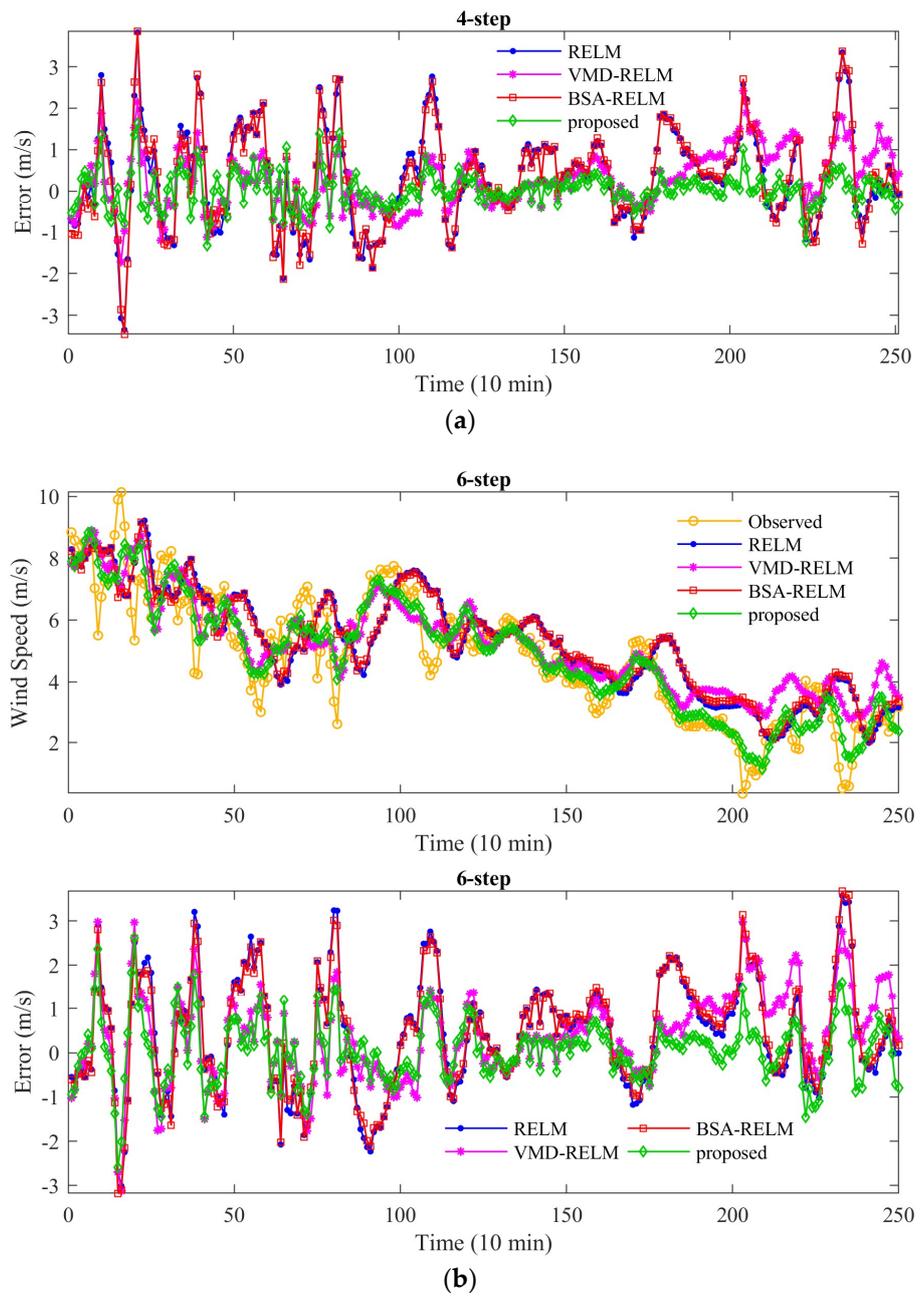

Figure 7. *Cont.*

Figure 7. 4- and 6-step forecasting results for Dataset A. (**a**) for 4-step and (**b**) for 6-step.
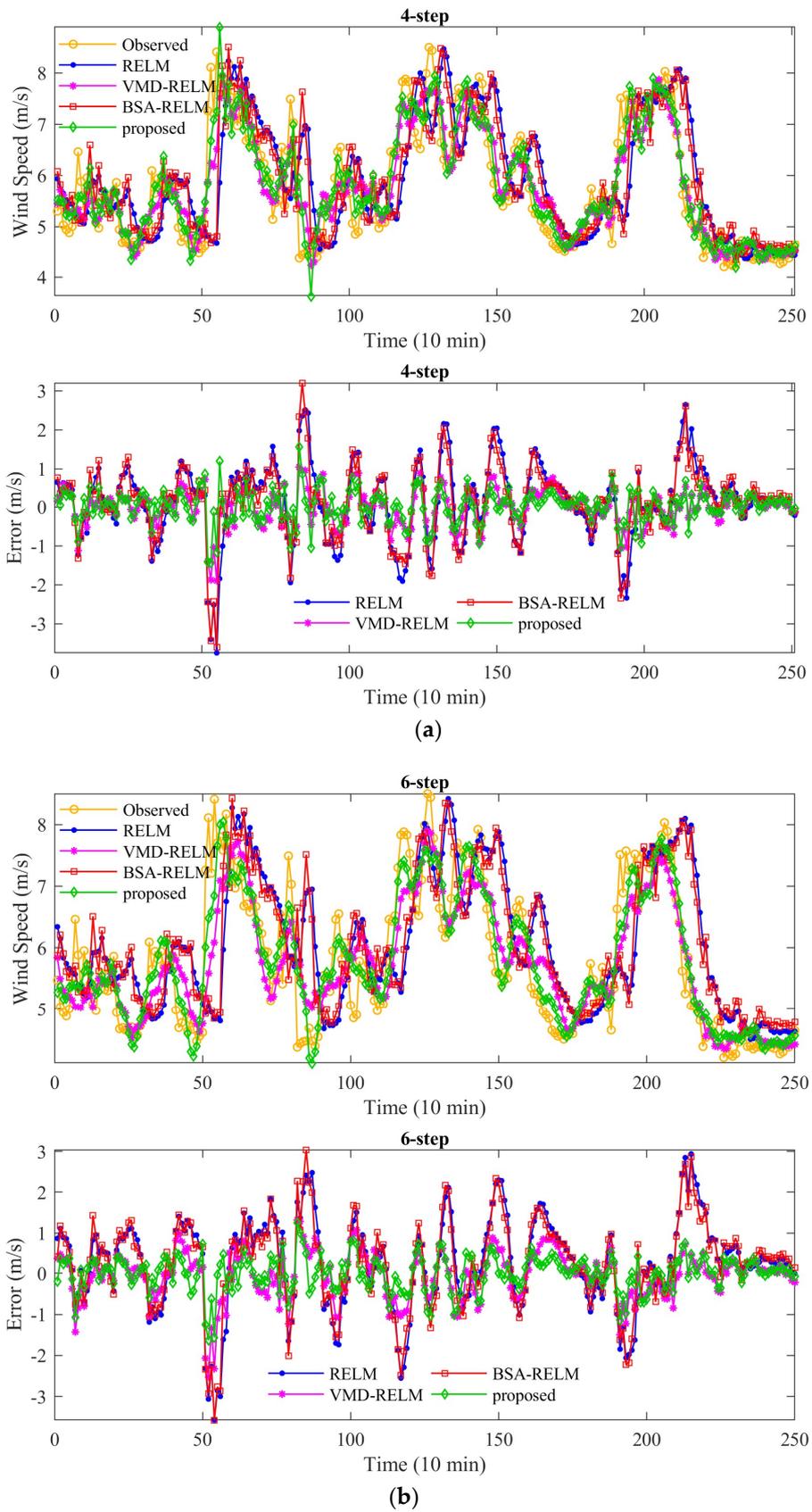
Figure 8. 4- and 6-step forecasting results for Dataset B. (**a**) for 4-step and (**b**) for 6-step.
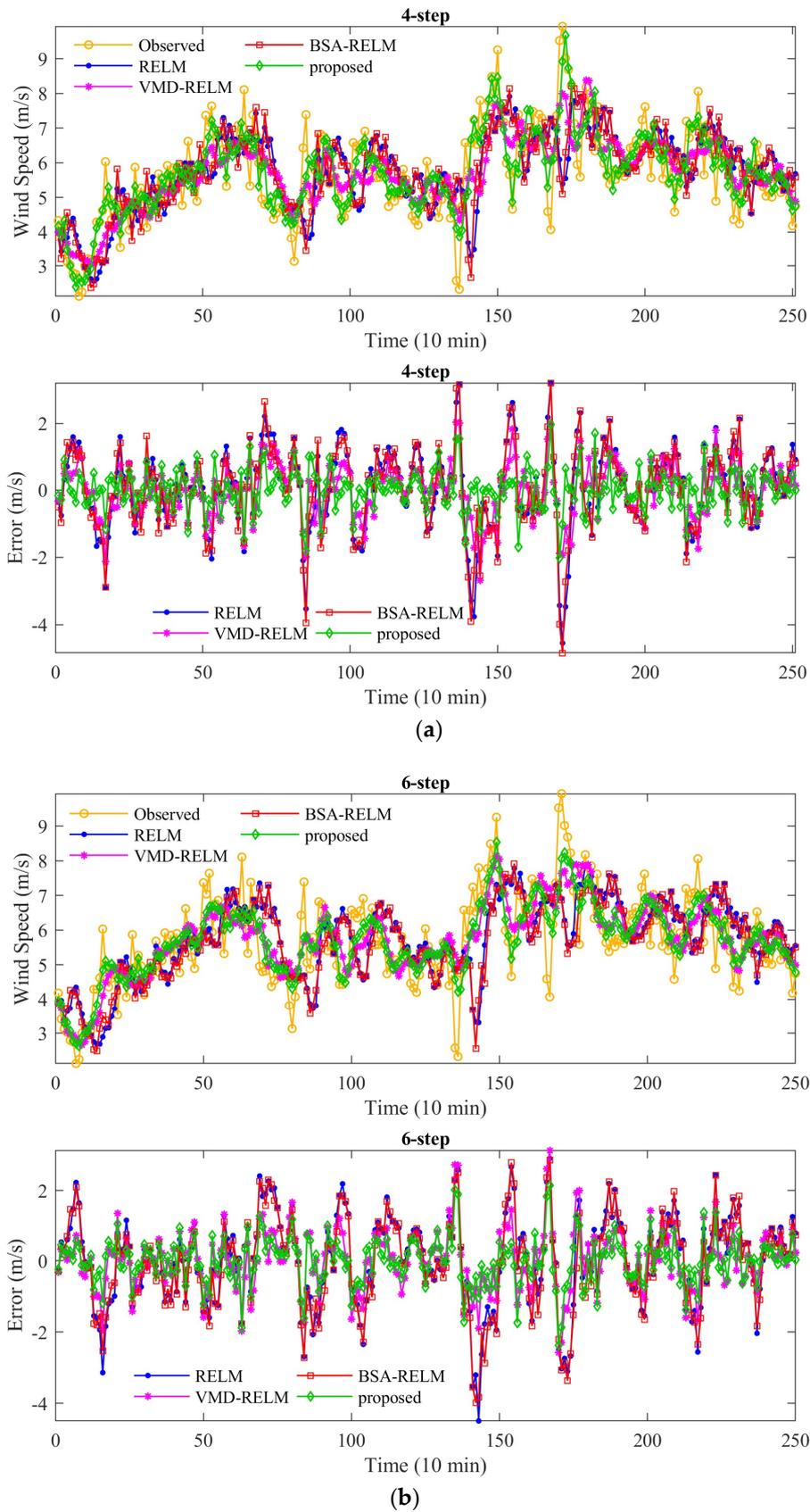
Figure 9. 4- and 6-step forecasting results for Dataset C. (**a**) for 4-step and (**b**) for 6-step.
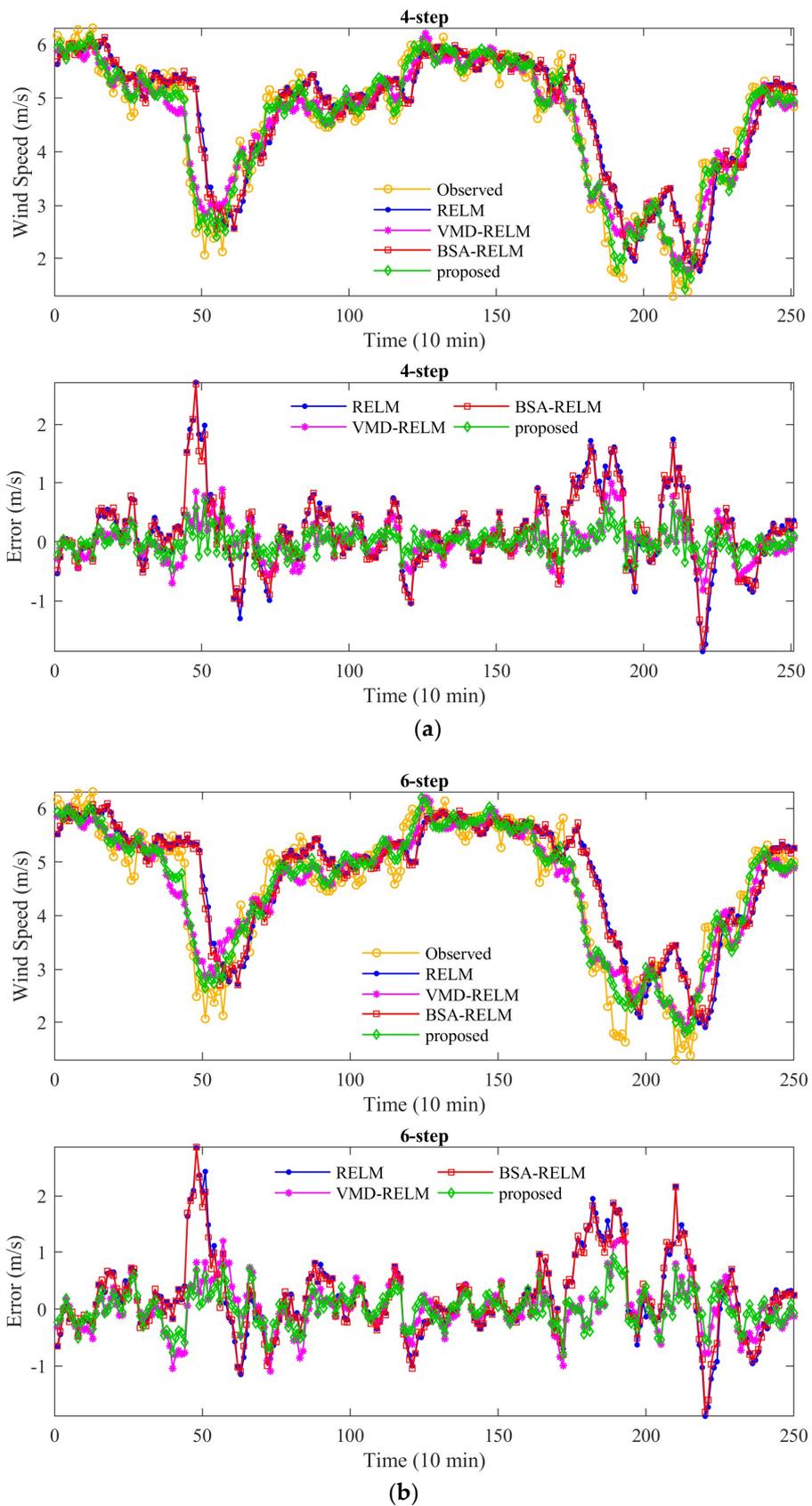
Figure 10. 4- and 6-step forecasting results for Dataset D. (**a**) for 4-step and (**b**) for 6-step.

The improved percentages of the proposed VMD-BSA-RELM model on the basis of RELM, VMD-RELM, and BSA-RELM for multi-step forecasting results are tabulated in Table 5. To visually view the improvements, the average improved percentages for different forecasting horizons in the last three rows of Table 5 are exhibited in Figure 11, where remarkable improvements of the proposed model over its competitors on all three metrics are revealed. The average improved ratios of all horizons for RELM, VMD-RELM, and BSA-ELM are 62.9%, 33.74%, and 61.63%, respectively. In summary, according to Tables 3 and 5, the proposed model produces a better result with a least 53.63% average improved ratios over three competitors with better performance among all benchmark models (RELM, VMD-RELM, and BSA-RELM) under all datasets and all horizons (1-, 2-, 4- and 6-step).
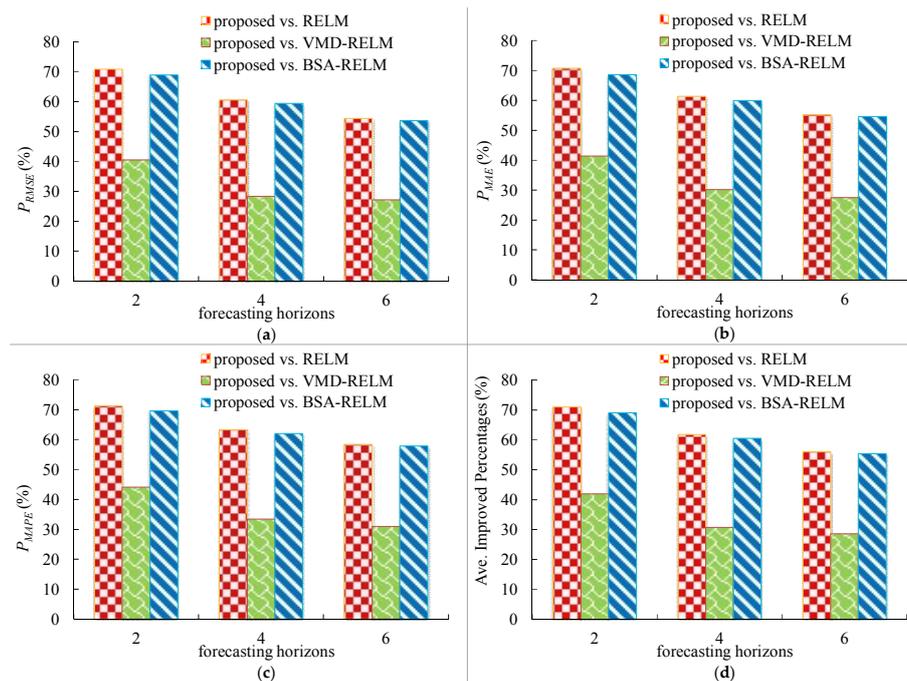

Figure 11. Average improved percentages under different horizons. (**a**) average P_{RMSE}; (**b**) average P_{MAE}; (**c**) average P_{MAPE}; and (**d**) average values of P_{RMSE}, P_{MAE}, and P_{MAPE}. (data from the last three rows of Table 5).

Table 5. Improved percentage of different error indices under multi-step forecasting.

Datasets	Forecasting Models	2-Step			4-Step			6-Step		
		P_{RMSE} (%)	P_{MAE} (%)	P_{MAPE} (%)	P_{RMSE} (%)	P_{MAE} (%)	P_{MAPE} (%)	P_{RMSE} (%)	P_{MAE} (%)	P_{MAPE} (%)
A	proposed vs. RELM	74.47	74.57	77.18	64.06	66.64	72.62	47.21	50.26	56.84
	proposed vs. VMD-RELM	59.23	61.38	71.05	43.19	48.05	62.90	33.22	36.14	49.57
	proposed vs. BSA-RELM	73.74	74.12	78.06	63.88	66.49	73.16	46.91	50.04	57.87
B	proposed vs. RELM	73.10	72.35	72.24	58.24	58.44	58.45	62.88	63.84	64.99
	proposed vs. VMD-RELM	36.96	39.28	39.51	10.52	11.15	9.97	28.34	27.39	26.36
	proposed vs. BSA-RELM	69.77	68.23	68.48	55.55	55.03	55.08	61.60	62.93	64.25
C	proposed vs. RELM	64.32	63.04	64.52	49.35	51.47	51.94	48.67	51.79	52.03
	proposed vs. VMD-RELM	37.04	34.46	35.32	23.61	27.11	26.82	23.29	24.61	25.21
	proposed vs. BSA-RELM	63.85	62.79	63.83	49.28	50.48	51.06	48.74	51.84	51.63
D	proposed vs. RELM	71.67	72.81	71.37	70.57	68.84	70.27	58.95	54.99	59.66
	proposed vs. VMD-RELM	28.83	30.58	30.46	36.12	34.60	33.98	24.01	21.79	22.67
	proposed vs. BSA-RELM	68.30	69.29	68.13	69.11	67.61	69.00	57.27	53.47	57.96
Average	proposed vs. RELM	70.43	69.98	72.15	59.28	60.40	65.13	53.58	54.84	57.78
	proposed vs. VMD-RELM	43.87	44.62	54.40	28.95	31.94	43.81	27.99	29.06	38.07
	proposed vs. BSA-RELM	68.76	68.29	71.47	58.27	59.11	64.57	52.81	54.28	57.66

5.3. Diebold-Mariano Tests and Computational Time

In this part, the Diebold-Mariano (DM) test [42] is applied to assess whether there are real differences between the proposed model and its competitors. The DM test results calculated by the square error loss function are tabulated in Table 6. It can be seen that the DM statistical values of the RELM, VMD-RELM, and BSA-RELM are more than the threshold value of the 1% significance level for all datasets and all forecasting horizons, which demonstrates the proposed model is superior to its rivals.

The average values of computational times of various step-ahead wind speed forecasting, with regard to Datasets A, B, C, and D, for all prediction models, are shown in Table 7. When compared with the other forecasting models, the proposed VMD-BSA-RELM model has higher time consumption due to the utilization of the optimization algorithm BSA, whereas its computational time is acceptable in real engineering application. These results have proven that the VMD-BSA-RELM model can provide more accurate wind speed forecasting results through sacrificing computational time within an admissible degree.

Table 6. Results for the DM test.

Datasets	Models	1-Step	2-Step	4-Step	6-Step
	RELM	8.1298 *	8.4312 *	10.0726 *	9.7100 *
Dataset A	VMD-RELM	9.2155 *	10.0883 *	8.2967 *	8.1698 *
	BSA-RELM	8.9351 *	8.7249 *	9.8394 *	9.7337 *
	RELM	6.0173 *	6.7680 *	8.2621 *	9.8667 *
Dataset B	VMD-RELM	3.1415 *	6.1912 *	2.3712 *	6.1402 *
	BSA-RELM	5.5741 *	5.9754 *	7.3984 *	9.5591 *
	RELM	8.4329 *	7.7476 *	7.5300 *	8.7539 *
Dataset C	VMD-RELM	6.2086 *	5.6829 *	5.2505 *	6.9023 *
	BSA-RELM	9.8363 *	8.4413 *	7.1543 *	8.8276 *
	RELM	7.7681 *	8.0019 *	7.6635 *	7.0102 *
Dataset D	VMD-RELM	3.3893 *	5.9483 *	7.2428 *	6.9873 *
	BSA-RELM	7.2513 *	8.0718 *	7.6262 *	6.8848 *

* indicates the 1% significance level.

Table 7. Computational time for different models. (unit: s).

Models	Dataset A	Dataset B	Dataset C	Dataset D	Average
RELM	0.2498	0.269	0.245	0.2515	0.253825
VMD-RELM	0.5305	0.5395	0.5188	0.506	0.5237
BSA-RELM	13.2803	12.8073	11.9297	12.3155	12.5832
VMD-BSA-RELM	40.87	40.2485	40.7218	39.2548	40.273775

6. Conclusions

Wind speed forecasting is a crucial part of wind energy generation. However, due to its inherent randomness, high non-linearity and non-stationarity, accurate wind speed forecasting is a very challengeable task. In this study, a new decomposition-optimization method called VMD-BSA-RELM is proposed for short-term wind speed forecasting. Original wind speed data is preprocessed by VMD into a group of relative stationary modes where regressions by RELM are executed. Suitable parameters of RELMs are determined by means of BSA. The efficacy of the VMD-BSA-RELM was tested against several benchmark models using several datasets under different forecasting horizons. The results indicate that the VMD-BSA-RELM model significantly outperforms the other models and sacrifices computational time with an acceptable degree in real applications. Additionally, quantitative analyses of the effects of decomposition and optimization techniques on the final improvement of forecasting

accuracy show that the decomposition approach, VMD, contributes more than the optimization method, BSA. The VMD-BSA-RELM combines VMD with BSA and thus take full advantages of both two methods. In conclusion, the proposed method is a powerful tool in short-term wind speed forecasting. Future work will focus on the investigation of the maximum forecasting horizons of the VMD-BSA-RELM model and introducing error correct techniques to improve its accuracy in large multi-step forecasting, especially in peak and valley parts.

Author Contributions: J.Z. and N.S. designed and performed the model simulation; N.S. wrote the draft of the paper; B.J. prepared the figures in the revised paper; T.P. polished the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

u_k	All modes of VMD
w_k	Frequencies of all modes
$f(t)$	Original signal
$\delta(t)$	Dirac distribution
α	The balancing parameter of the data-fidelity constraint
$\hat{u}_k^{n+1}$	Fourier transform of u_k^{n+1}
$\hat{u}_i(\omega)$	Fourier transform of $u_i(t)$
$\hat{f}(\omega)$	Fourier transform of $f(t)$
$\hat{\lambda}(\omega)$	Fourier transform of $u_i(t)$
τ	Time-step of the dual ascent
ε	Tolerance of convergence criterion
$g_i(x)$	Activation function
w_i	The input weight vector
H	The hidden layer output matrix
$H^\dagger$	Moore-Penrose generalized inverse matrix of H
(x_t, y_t)	Training samples
M	Total number of training samples
L	hidden nodes
N	Population size
D	Individual dimensionality
a, b	Random numbers distributed uniformly from 0 to 1
$OldP$	Historical population
T	Trial population
P	Current population
$mixrate$	Mix rate parameter of BSA
h	Forecasting horizon
d	Suitable lag time

References

1. Global Wind Energy Council (GWEC). *Global Wind Report 2016-Annual Market Update*; GWEC: Brussels, Belgium, 2016.
2. Soman, S.S.; Zareipour, H.; Malik, O.; Mandal, P. A review of wind power and wind speed forecasting methods with different time horizons. In Proceedings of the 2010 North American Power Symposium, Arlington, TX, USA, 26–28 September 2010; pp. 1–8.
3. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [CrossRef]

4. Sun, N.; Zhou, J.Z.; Zhu, S.; Wei, L.I.; Peng, T. Application of Hybrid Models Based on Wavelet Analysis and Two Different Neural Networks in Prediction of Monthly Runoff. *Water Resour. Power* **2018**. [CrossRef]

5. Sun, N.; Zhou, J. Non-stationary runoff hybrid forecasting model based on regularized extreme learning machine. *J. Hydroel. Eng.* **2018**, *1*, 1–9.

6. Cincotti, S.; Gallo, G.; Ponta, L.; Raberto, M. Modeling and forecasting of electricity spot-prices: Computational intelligence vs. classical econometrics. *Ai Commun.* **2014**, *27*, 301–314.

7. Lydia, M.; Kumar, S.S.; Selvakumar, A.I.; Kumar, G.E.P. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Convers. Manag.* **2016**, *112*, 115–124. [CrossRef]

8. Han, Q.; Meng, F.; Hu, T.; Chu, F. Non-parametric hybrid models for wind speed forecasting. *Energy Convers. Manag.* **2017**, *148*, 554–568. [CrossRef]

9. Wang, J.; Hu, J. A robust combination approach for short-term wind speed forecasting and analysis–Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy* **2015**, *93*, 41–56.

10. Liu, H.; Tian, H.; Liang, X.; Li, Y. New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks. *Renew. Energy* **2015**, *83*, 1066–1075. [CrossRef]

11. Ren, C.; An, N.; Wang, J.; Li, L.; Hu, B.; Shang, D. Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting. *Knowl.-Based Syst.* **2014**, *56*, 226–239. [CrossRef]

12. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [CrossRef]

13. Wang, J.; Heng, J.; Xiao, L.; Wang, C. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. *Energy* **2017**, *125*, 591–613. [CrossRef]

14. Wang, J.; Yang, W.; Du, P.; Li, Y. Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. *Energy* **2018**, *148*, 59–78. [CrossRef]

15. Yu, C.; Li, Y.; Zhang, M. An improved Wavelet Transform using Singular Spectrum Analysis for wind speed forecasting based on Elman Neural Network. *Energy Convers. Manag.* **2017**, *148*, 895–904. [CrossRef]

16. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]

17. MartíNez-MartíNez, J.M.; Escandell-Montero, P.; Soria-Olivas, E.; MartíN-Guerrero, J.D.; Magdalena-Benedito, R.; GóMez-Sanchis, J. Regularized extreme learning machine for regression problems. *Neurocomputing* **2011**, *74*, 3716–3721. [CrossRef]

18. Chen, K.; Lv, Q.; Lu, Y.; Dou, Y. Robust regularized extreme learning machine for regression using iteratively reweighted least squares. *Neurocomputing* **2017**, *230*, 345–358. [CrossRef]

19. Huang, N.; Yuan, C.; Cai, G.; Xing, E. Hybrid Short Term Wind Speed Forecasting Using Variational Mode Decomposition and a Weighted Regularized Extreme Learning Machine. *Energies* **2016**, *9*, 989. [CrossRef]

20. Shao, H.; Deng, X.; Cui, F. Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of East China. *Iet Gener. Transm. Distrib.* **2016**, *10*, 2585–2592. [CrossRef]

21. Liu, D.; Niu, D.; Wang, H.; Fan, L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew. Energy* **2014**, *62*, 592–597. [CrossRef]

22. Wang, Y.; Wu, L. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. *Energy* **2016**, *112*, 208–220. [CrossRef]

23. Guo, Z.; Zhao, W.; Lu, H.; Wang, J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renew. Energy* **2012**, *37*, 241–249. [CrossRef]

24. Yu, C.; Li, Y.; Zhang, M. Comparative study on three new hybrid models using Elman Neural Network and Empirical Mode Decomposition based technologies improved by Singular Spectrum Analysis for hour-ahead wind speed forecasting. *Energy Convers. Manag.* **2017**, *147*, 75–85. [CrossRef]

25. Abdoos, A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **2016**, *203*, 111–120. [CrossRef]

26. Zhang, Y.; Liu, K.; Qin, L.; An, X. Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods. *Energy Convers. Manag.* **2016**, *112*, 208–219. [CrossRef]

27. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Transac. Signal Process.* **2014**, *62*, 531–544. [CrossRef]

28. Gao, Y.; Qu, C.; Zhang, K. A Hybrid Method Based on Singular Spectrum Analysis, Firefly Algorithm, and BP Neural Network for Short-Term Wind Speed Forecasting. *Energies* **2016**, *9*, 757. [CrossRef]

29. Yin, H.; Dong, Z.; Chen, Y.; Ge, J.; Lai, L.L.; Vaccaro, A.; Meng, A. An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. *Energy Convers. Manag.* **2017**, *150*, 108–121. [CrossRef]

30. Yuan, X.; Chen, C.; Yuan, Y.; Huang, Y.; Tan, Q. Short-term wind power prediction based on LSSVM–GSA model. *Energy Convers. Manag.* **2015**, *101*, 393–401. [CrossRef]

31. Osório, G.; Matias, J.; Catalão, J. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. *Renew. Energy* **2015**, *75*, 301–307. [CrossRef]

32. Civicioglu, P. Backtracking Search Optimization Algorithm for numerical optimization problems. *Appl. Math. Comput.* **2013**, *219*, 8121–8144. [CrossRef]

33. Zhang, C.; Zhou, J.; Li, C.; Fu, W.; Peng, T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *143*, 360–376. [CrossRef]

34. Taslimi-Renani, E.; Modiri-Delshad, M.; Elias, M.F.M.; Rahim, N.A. Development of an enhanced parametric model for wind turbine power curve. *Appl. Energy* **2016**, *177*, 544–552. [CrossRef]

35. Renani, E.T.; Elias, M.F.M.; Rahim, N.A. Using data-driven approach for wind power prediction: A comparative study. *Energy Convers. Manag.* **2016**, *118*, 193–203. [CrossRef]

36. Xiao, L.; Shao, W.; Yu, M.; Ma, J.; Jin, C. Research and application of a combined model based on multi-objective optimization for electrical load forecasting. *Energy* **2017**, *119*, 1057–1074. [CrossRef]

37. Salcedo-Sanz, S.; Pastor-Sánchez, A.; Prieto, L.; Blanco-Aguilera, A.; García-Herrera, R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–Extreme learning machine approach. *Energy Convers. Manag.* **2014**, *87*, 10–18. [CrossRef]

38. Wei, S.; Liu, M. Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China. *Energy Convers. Manag.* **2016**, *114*, 197–208.

39. Wang, J.Z.; Wang, Y.; Jiang, P. The study and application of a novel hybrid forecasting model–A case study of wind speed forecasting in China. *Appl. Energy* **2015**, *143*, 472–488. [CrossRef]

40. Chen, L.; Singh, V.P. Entropy-based derivation of generalized distributions for hydrometeorological frequency analysis. *J. Hydrol.* **2018**, *557*, 699–712. [CrossRef]

41. Chen, L.; Singh, V.P.; Huang, K. Bayesian technique for the selection of probability distributions for frequency analysis of hydrometeorological extremes. *Entropy* **2018**, *20*, 117. [CrossRef]

42. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [CrossRef]